



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS VII – GOVERNADOR ANTÔNIO MARIZ
CENTRO DE CIÊNCIAS EXATAS E SOCIAIS APLICADAS
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

KEILA LUCAS DOS SANTOS

**MAPEAMENTO DOS CASOS DA DOENÇA DENGUE NO ESTADO DA PARAÍBA:
IDENTIFICAÇÃO DAS ÁREAS DE RISCO**

**PATOS - PB
2019**

KEILA LUCAS DOS SANTOS

**MAPEAMENTO DOS CASOS DA DOENÇA DENGUE NO ESTADO DA PARAÍBA:
IDENTIFICAÇÃO DAS ÁREAS DE RISCO**

Trabalho de Conclusão de Curso apresentado ao Curso Bacharelado em Ciência da Computação da Universidade Estadual da Paraíba, em cumprimento para obtenção do grau de bacharelado em Ciência da Computação.

Orientador: Prof. Me. Pablo Roberto Fernandes de Oliveira.

**PATOS - PB
2019**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

S237m Santos, Keila Lucas dos.
Mapeamento dos casos da doença dengue no estado da Paraíba [manuscrito] : identificação das áreas de risco / Keila Lucas dos Santos. - 2019.
49 p. : il. colorido.
Digitado.
Trabalho de Conclusão de Curso (Graduação em Computação) - Universidade Estadual da Paraíba, Centro de Ciências Exatas e Sociais Aplicadas, 2019.
"Orientação : Prof. Me. Pablo Roberto Fernandes de Oliveira, Coordenação do Curso de Computação - CCEA."
1. Dengue. 2. Ciência de dados. 3. Linguagem R. I. Título
21. ed. CDD 004

Keila Lucas dos Santos

**MAPEAMENTO DOS CASOS DA DOENÇA DENGUE NO ESTADO DA PARAÍBA:
IDENTIFICAÇÃO DAS ÁREAS DE RISCO**

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Ciências da Computação da Universidade Estadual da Paraíba, em cumprimento à exigência para obtenção do grau de Bacharel em Ciência da Computação.

Aprovado em 28/11/2019

BANCA EXAMINADORA

Pablo Roberto Fernandes de Oliveira

Prof. Me. Pablo Roberto F. de Oliveira
(Orientador)

Fábio Júnior Francisco da Silva

Prof. Esp. Fábio Júnior F. da Silva
(Examinador)

Wellington C. Araújo

Prof. Dr. Wellington Candeia de Araújo
(Examinador)

Aos meus pais, por todos os ensinamentos de vida.

AGRADECIMENTOS

Aos meus pais, Inês Lucas e José Nunes, pelo amor, incentivo e apoio incondicional.

Aos meus irmãos, Raniery, Yuri e Uriel pelo companheirismo de sempre.

Ao meu noivo Fagundes, por todo carinho e incentivo, pelas loucuras e risadas.

Obrigada principalmente por me dar suporte quando fracassei.

Aos meus amigos da turma de Computação 2015.1, vocês são tão incríveis quanto loucos, ou melhor, vocês são muito “zueiros”. Conhecê-los e viver a graduação com vocês foi maravilhoso.

Agradeço ao meu professor e orientador Pablo Roberto, pela orientação, apoio e confiança.

Agradeço ao professor Wellington Candeia, por todas as oportunidades permitidas durante a iniciação científica.

Agradeço também aos muitos professores que ao longo dos anos contribuíram para o meu progresso profissional e pessoal, levarei muito de vocês comigo. Especialmente aqueles que reconheço como verdadeiros amigos e possuem todo meu respeito e admiração.

Agradeço imensamente a UEPB, por todo suporte educacional e pela chance de evoluir meus conhecimentos.

“Nós só podemos ver um pouco do futuro, mas o suficiente para perceber que há muito a fazer.”

(Alan Turing)

RESUMO

A doença dengue tornou-se, nas últimas décadas, uma das maiores preocupações no contexto da saúde pública. Surto recorrentes, com diversos fatores de agravo despertam estado de alerta em várias regiões do Brasil, nas quais milhares de pessoas são afetadas pela doença. Este trabalho teve como objetivo realizar o mapeamento dos casos de dengue no Estado da Paraíba, no período de 2014 a 2018 por meio da análise exploratória de dados referentes às internações hospitalares com diagnóstico de dengue, a fim de identificar os municípios mais vulneráveis e verificar os número de internações em relação a densidade demográfica, volume de chuvas e aos focos do mosquito *Aedes aegypti*. A análise dos dados das Autorizações de Internações Hospitalares (AIH) foi realizada utilizando os pacotes *read.dbc*, *stringr* e *lubridate* da linguagem R, assim como funções estatísticas de frequência, soma, máximo e mínimo para distribuição dos dados, e o software de processamento geográfico QGIS 3.4 para o desenvolvimento de mapas graduados. Como resultado, verificou-se que os municípios de Campina Grande, Araruna, Uiraúna e João Pessoa foram os mais afetados pela dengue no período de 2014 a 2018, destacando-se que a maior incidência de casos na Paraíba tende a ocorrer no período entre os meses de março e junho. Além disso, foi possível observar no contexto paraibano que a densidade demográfica não apresenta relação significativa com os valores de internações observados.

Palavras-Chave: Dengue. Análise Exploratória de Dados. Mapeamento. Ciência de Dados.

ABSTRACT

Dengue disease has in recent decades become one of the major concerns in the context of public health. Recurring outbreaks with various grievance factors arouse alertness in several regions of Brazil, where thousands of people are affected by the disease. This study aimed to map dengue cases in the state of Paraíba, from 2014 to 2018 through exploratory analysis of data regarding hospitalizations diagnosed with dengue, in order to identify the most vulnerable municipalities and verify the the number of hospitalizations in relation to demographic density, rain volume and spotlights of *Aedes aegypti* mosquito. Data analysis of Hospitalization Authorization Form (Known in Brazil by the Portuguese acronym, AIH) was performed using the R language read.dbc, stringr and lubidate packages, as well as statistical functions of frequency, sum, maximum and minimum for data distribution and the processing software geographic QGIS 3.4 for the development of graded maps. As a result, it was found that the municipalities of Campina Grande, Araruna, Uiraúna and João Pessoa were the most affected by dengue in the period from 2014 to 2018, highlighting that the highest incidence of cases in Paraíba tends to occur between March and July. In addition, it was possible to observe in the Paraíba context that demographic density has no significant relationship with the values of hospitalizations observed.

Keywords: Dengue. Exploratory Data Analysis. Mapping. Data Science.

LISTA DE ABREVIATURAS

| | | | |
|-----|------------------------|------|--------------------------|
| AN | ARARUNA | JP | JOÃO PESSOA |
| AR | AREIAL | LS | LAGOA SECA |
| ARO | AROEIRAS | LT | LASTRO |
| BA | BARAÚNA | MT | MONTEIRO |
| BBC | BELÉM DO BREJO DO CRUZ | PO | POMBAL |
| BC | BREJO DO CRUZ | PC | POCINHOS |
| BDS | BREJO DOS SANTOS | PD | POÇO DANTAS |
| BE | BAYEUX | PI | PRINCESA ISABEL |
| BN | BANANEIRAS | PL | PEDRA LAVRADA |
| BO | BOQUEIRÃO | PN | PIANCÓ |
| BS | BOM SUCESSO | PT | PATOS |
| BSF | BONITO DE SANTA FÉ | QE | QUEIMADAS |
| CA | CAAPORÃ | RC | RIACHO DOS CAVALOS |
| CAS | CASSERENGUE | RDB | RIACHÃO DO BACAMARTE |
| CB | CUBATI | SB | SÃO BENTO |
| CG | CAMPINA GRANDE | SBA | SERRA BRANCA |
| CJ | CAJAZEIRAS | SC | SANTA CRUZ |
| CO | COREMAS | SJP | SÃO JOSÉ DE PIRANHAS |
| CR | CATOLÉ DO ROCHA | SJRP | SÃO JOÃO DO RIO DO PEIXE |
| CT | CATURITÉ | SL | SOLÂNEA |
| CTE | CUITÉ | SM | SÃO MAMEDE |
| FG | FAGUNDES | SR | SANTA RITA |
| GB | GUARABIRA | SS | SOUSA |
| IG | INGÁ | SVS | SÃO VICENTE DO SERIDÓ |
| JE | JERICÓ | UN | UIRAÚNA |

SUMÁRIO

| | | |
|------------|---|-----------|
| 1 | INTRODUÇÃO..... | 10 |
| 2 | REFERENCIAL TEÓRICO..... | 13 |
| 2.1 | Dengue..... | 13 |
| 2.2 | Dados AIH/SUS..... | 16 |
| 2.2.1 | Descrição da AIH..... | 17 |
| 2.3 | Ciência de Dados..... | 18 |
| 2.3.1 | Análise Exploratória de Dados..... | 20 |
| 2.4 | Série Temporal..... | 21 |
| 2.5 | Linguagem R..... | 22 |
| 2.6 | Trabalhos Relacionados..... | 24 |
| 3 | METODOLOGIA..... | 26 |
| 3.1 | Coleta de dados..... | 28 |
| 3.1.1 | Tratamento dos dados..... | 29 |
| 3.2 | Tabulação dos dados..... | 30 |
| 4 | ANÁLISE E DISCUSSÃO DOS RESULTADOS..... | 31 |
| 4.1 | Casos de dengue no Estado da Paraíba..... | 31 |
| 4.2 | Mapeamento das áreas de risco no Estado da Paraíba..... | 35 |
| 4.3 | Análise do número de internações em relação à densidade demográfica, volume de chuvas e aos focos do <i>Aedes aegypti</i>..... | 39 |
| 5 | CONSIDERAÇÕES FINAIS..... | 42 |
| 5.1 | Sumário da Pesquisa..... | 42 |
| 5.2 | Limitações da Pesquisa..... | 44 |
| 5.3 | Trabalhos Futuros..... | 44 |
| | REFERÊNCIAS..... | 46 |

1 INTRODUÇÃO

Na América do Sul foi observado que nas últimas três décadas ocorreu um aumento significativo dos casos de arboviroses, que são as doenças transmitidas por mosquitos. Em destaque para o *Aedes aegypti* que é o principal vetor de doenças como a Febre Dengue (clássica ou hemorrágica), Febre Zika e Febre Chikungunya (FOGAÇA; MENDONÇA, 2017).

O controle das epidemias ocasionadas por este mosquito é dificultado, principalmente pelos fatores ambientais e sociais presentes nos cenários vulneráveis. Dentre os trabalhos de combate a proliferação do mosquito da dengue, se destacam as atividades realizadas pelos Agentes de Combate às Endemias (ACE), que são responsáveis por utilizar as medidas de controle químico, biológico, manejo ambiental e outras ações de manejo integrado de vetores (BRASIL, 2018).

Apesar dos esforços dedicados por esses profissionais, novos surtos são recorrentes em diversas regiões do País. Conforme Souza et al. (2018), as dificuldades de acesso aos imóveis e terrenos baldios - que geralmente apresentam muito mato e lixo acumulado - são alguns dos principais fatores que inviabilizam em parte, o trabalho dos ACE.

Segundo o Ministério da Saúde, o número de casos prováveis de dengue no Brasil foi mais que duplicado em janeiro de 2019, quando comparado ao mesmo período de 2018. Diante desta situação, inúmeras ações precisam ser desenvolvidas, principalmente para dar suporte aos indivíduos infectados pela doença (BRASIL, 2019d).

Esse aumento de casos da doença desperta questionamentos sobre quais são os fatores agravantes ao desenvolvimento do foco do mosquito e quais são as áreas mais vulneráveis em determinada região. Na cidade de Patos-PB, através da pesquisa de Oliveira (2018), foi possível identificar que 4 áreas correspondentes aos bairros Monte Castelo, Jatobá, Alto da Tubiba e Conjunto Nova conquista (área 1), Bivar Olinto, Santo Antônio e Maternidade (área 2), Centro, Liberdade e Morro (área 3), Salgadinho, São Sebastião e Sete casas (área 4) apresentaram alto risco de surto da dengue, devido a existência de pontos com lixo a céu aberto, esgoto a céu aberto e predominância de população com renda familiar de um salário mínimo.

A identificação dessas áreas vulneráveis na cidade de Patos auxiliou no direcionamento das ações de combate à dengue. Com isso, é possível ressaltar a importância de desenvolver em outros estudos a expansão deste tipo de análise para outros municípios. No entanto, desenvolver um trabalho desta natureza requer a utilização de técnicas computacionais para realizar o processamento dos dados relacionados aos casos de dengue.

No que diz respeito aos dados da dengue, o Ministério da Saúde divulga através do portal do Departamento de Informática do Sistema Único de Saúde (DATASUS), diversos dados referentes a dengue no Brasil e em suas unidades federativas. Esses dados, apesar de serem abertos, ou seja, podem ser acessados pela população - principalmente para o desenvolvimento de pesquisas - são dispostos em enormes bases de armazenamento. Logo, realizar uma análise sobre eles sem suporte computacional é uma atividade inviável.

Considerando este fator, a aplicação de técnicas da Ciência de Dados é extremamente útil para reconhecer padrões e descobrir conhecimento através da manipulação de grandes quantidades de dados (PORTO; ZIVIANI, 2014). Também é importante ressaltar que diversas pesquisas atualmente, como Dharmawardana et al. (2017) e Mathulamuthu et al. (2016), aplicam a metodologia da Ciência de Dados, com foco nos dados de arboviroses com objetivo de desenvolver modelos preditivos através das técnicas de Aprendizagem de Máquina. Além disso, esses modelos preditivos são desenvolvidos com o uso dos algoritmos de inteligência artificial (IA) capazes de identificar a incidência de casos de dengue em uma determinada região geográfica (DAVI et al., 2019).

Diante da problemática apresentada, este trabalho tem o objetivo de mapear os casos da doença dengue no Estado da Paraíba, identificando os locais de vulnerabilidade através da análise exploratória dos dados abertos presentes nos formulários das Autorizações de Internações Hospitalares (AIH) no período de 2014 a 2018 obtidos no portal DATASUS.

A pesquisa possui os seguintes objetivos específicos:

(OBE1) Identificar quais foram os municípios com o maior número de internações hospitalares;

(OBE2) Verificar o comportamento da doença no Estado da Paraíba ao longo dos meses em cada ano de 2014 a 2018;

(OBE3) Verificar a relação entre o número de internações e as algumas variáveis como a densidade demográfica dos municípios, o volume de chuva e os focos do mosquito vetor.

Em relação aos formulários de AIH, as discussões sobre importância dos dados contidos nestes documentos foram levantadas nas reuniões de desenvolvimento do projeto de Iniciação Científica intitulado “Aprendizado de máquina para modelagem de doenças causadas por arbovírus”, na Universidade Estadual da Paraíba, Campus VII - Patos.

A decisão de trabalhar com AIH é decorrente da confiabilidade presente nos dados inseridos nos formulários. Essas autorizações devem ser preenchidas corretamente para que os hospitais públicos e privados vinculados ao Sistema Único de Saúde (SUS) possam receber o reembolso da assistência prestada aos pacientes. Assim, todos os registros são auditados com o

objetivo de evitar erros e inconsistências nos dados enviados (VERAS; MARTINS, 1994; CUNHA,2013).

Portanto, com os resultados obtidos através da aplicação de técnicas da Ciência de Dados na manipulação de AIH, esta pesquisa busca auxiliar na definição de diretrizes para prevenção e controle das epidemias da dengue na Paraíba, visando combater a propagação de surtos da doença.

2 REFERENCIAL TEÓRICO

2.1 Dengue

As arboviroses são doenças transmitidas por mosquitos, entre eles o *Aedes aegypti*, que é o mosquito vetor da dengue. Os surtos decorrentes da transmissão dessas doenças são uma preocupação mundial de saúde pública. De acordo com a atualização epidemiológica da Organização Pan-Americana da Saúde (OPAS), em 2018 foram notificados 560.586 casos de dengue nas Américas, destes casos, 3.535 foram classificados como dengue grave (dengue hemorrágica) e ocorreram 336 mortes. O relatório inicial do ano de 2019, em suas primeiras seis semanas, apresentou quase 100 mil casos de dengue notificados, com 632 indicações de dengue grave e 28 óbitos. (DAVI et al., 2019; ONU, 2019)

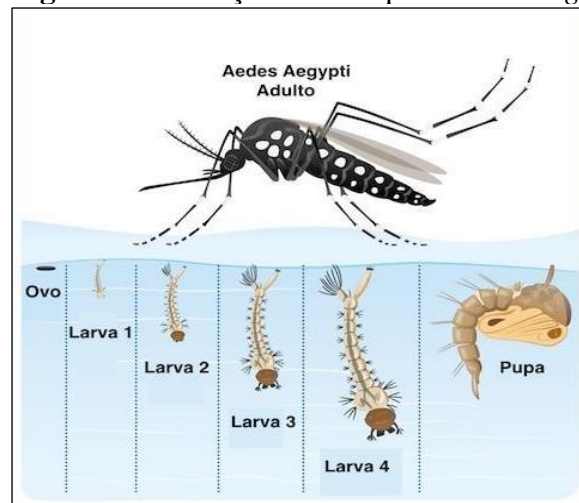
A dengue é uma doença infecciosa febril aguda causada por vírus que pertence à família *Flaviviridae*, do gênero *Flavivirus*. Segundo Ramalho et al. (2018), o vírus da dengue apresenta quatro sorotipos do Dengue Vírus (DENV), com alta variabilidade genotípica. Através de estudos de sequenciamento genômico identificou-se cinco genótipos para DENV-1, seis genótipos para DENV-2, cinco genótipos para DENV-3 e quatro genótipos para DENV-4.

A contaminação viral ocorre quando a fêmea do mosquito *Aedes aegypti*, após picar uma pessoa infectada, transmite o vírus para outras pessoas (BRASIL, 2019b). Este fator de transmissão quando relacionado à movimentação de pessoas em centros urbanos impulsiona os surtos da doença a progredirem para outras regiões vulneráveis, conforme é apresentado no estudo de Dharmawardana et al. (2017).

Outro fator conveniente para a proliferação do mosquito e consequentemente a recorrência da doença é o clima. Regiões tropicais e subtropicais do planeta apresentam condições favoráveis ao desenvolvimento do mosquito (NASCIMENTO et al., 2015). Segundo Mathulamuthu et al. (2016), os mosquitos vetores são sensíveis à temperatura, chuva e umidade, e por isso mudanças no clima são tão relevantes para as incidências da dengue. Além disso, os ambientes quentes e úmidos fornecem ao óvulo do mosquito um estado adequado para sua sobrevivência por mais de um ano até a eclosão na água. Com isso, o trabalho de verificação do foco do mosquito em recipientes com água parada é fundamental para controlar novos surtos da dengue.

Na Figura 1 apresenta-se as 7 fases de evolução do mosquito *Aedes aegypti*. Progredindo do ovo para quatro tamanhos de larva, pupa e finalmente atingindo a fase adulta.

Figura 1 - Evolução do mosquito *Aedes aegypti*



Fonte: CDC (2018).

O Ministério da Saúde alerta que o mosquito sobrevive e pode transmitir arboviroses em qualquer época do ano, porém nos meses mais quentes e chuvosos ocorre maior disponibilidade de pequenas ou médias acumulações de água favorecendo o desenvolvimento do vetor (BRASIL, 2019b).

Todas as faixas etárias são igualmente suscetíveis, porém as pessoas mais vulneráveis são as crianças e idosos com mais de 65 anos. Os idosos têm maior risco de desenvolver dengue grave porque o vírus pode interagir com doenças pré-existentes e as complicações decorrentes que podem levar à morte. Indivíduos com doenças crônicas, como diabetes e hipertensão arterial também são vulneráveis para desenvolver um quadro grave da doença (BRASIL, 2019b).

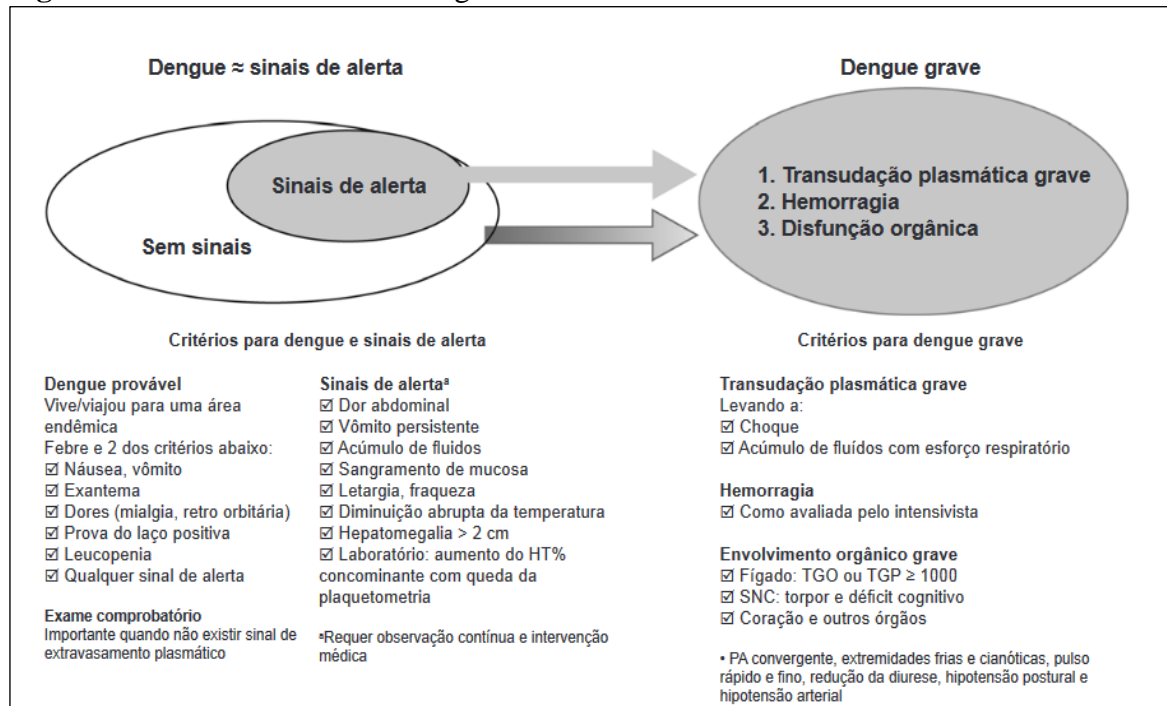
Quando a dengue progride para estágios mais graves, os indivíduos infectados passam a correr risco de vida devido o surgimento de complicações, incluindo vazamento vascular, falência de órgãos e síndrome do choque. As sequelas decorrentes da infecção viral podem levar ao desenvolvimento síndrome de Guillain-Barré, encefalite e outras complicações neurológicas (DAVI et al., 2019; BRASIL, 2019b).

O manejo dos pacientes com sinais de alarme ou nos casos graves deve ser estruturado de forma que o tratamento seja prioritário e apropriado de acordo com a classificação de risco. Dessa forma, os procedimentos de constante reavaliação e observação são essenciais para que as intervenções sejam corretas com objetivo de evitar óbitos (BRASIL, 2016).

Em decorrência dos sinais de alerta (ver Figura 2), torna-se indispensável a execução de procedimentos para a confirmação diagnóstica dos pacientes que apresentam as formas graves

da doença, gestantes, crianças ou aqueles indivíduos que demonstram sintomas atípicos (VERDAL et al., 2011).

Figura 2 - Sinais de Alerta da Dengue.



Fonte: Verdal et al. (2011).

Após a percepção dos sinais de alertas, os pacientes identificados neste quadro clínico deverão ser internados em unidades de observação para que os exames laboratoriais sejam realizados a fim de se obter a confirmação da dengue. Segundo Verdal et al. (2011), no período entre o primeiro e o quinto dia, por ainda não haver conversão sorológica, é possível diagnosticar a infecção através de isolamento viral em cultura de células. A partir do sexto dia utiliza-se o método ELISA-IgM para o diagnóstico da dengue.

Caso ocorra deterioração clínico-laboratorial, os pacientes devem ser transferidos para Unidade de Terapia Intensiva (UTI). Alguns pacientes que apresentam patologias como: Obesidade, diabetes mellitus, doença grave do sistema cardiovascular, asma brônquica e doença pulmonar entre outras, pessoas com menos de 15 ou mais de 60 anos e mulheres grávidas requerem atenção especial devido a possibilidade de agravamento das condições pré-existentes (VERDAL et al., 2011).

Todos os pacientes que precisam ser internados nos hospitais da rede do SUS, são reconhecidos através da emissão de um laudo que solicita a autorização para que internação seja realizada. Esta autorização é reconhecida como AIH, que se trata do documento necessário para identificar o paciente e os serviços prestados sob o regime de internação. Este documento

deve ser corretamente preenchido pelo profissional médico, que descreve as condições que justificam a internação e o diagnóstico inicial (BRASIL, 2019a).

A AIH é composta por um conjunto de variáveis¹, das quais destaca-se: **Código do Hospital, Identificação do paciente, Recursos pagos, Diagnóstico principal** (motivo da internação). Essas autorizações são manipuladas pelos gestores e prestadores de serviços do SUS, responsáveis por inserir corretamente todos os dados referentes às despesas dos procedimentos médico-hospitalares realizados nos pacientes (BRASIL, 2017).

2.2 Dados AIH/SUS

O Ministério da Saúde implantou o Sistema SIH/SUS através da Portaria GM/MS n.º 896/1990. Este sistema processa os registros de atendimentos realizados na rede SUS, que são coletados para compor o Banco de Dados Nacional (BRASIL, 2017).

Os gestores hospitalares são responsáveis por coordenar, avaliar e auditar estes registros, que são repassados e integrados ao acervo de dados do sistema de saúde pelo DATASUS.

Por meio do DATASUS, a população dispõe do acesso aos dados de saúde, que podem ser utilizadas como suporte nas tomadas de decisão, decorrente de análises objetivas nas diversas áreas da rede de atendimento público. Dentre os dados acessíveis estão os registros referentes às Estatísticas Vitais - Mortalidade e Nascidos Vivos, Epidemiológicas e Morbidade, Indicadores de Saúde, Demográficas e Socioeconômicas, AIH, entre outros².

Os dados presentes nas AIH são registrados pelos profissionais prestadores de serviços do SUS, com o objetivo de identificar o paciente e as despesas decorrentes dos procedimentos médicos realizados enquanto esta pessoa estiver em estado de internação (BRASIL, 2019a). É através da emissão deste registro que o SUS ampara o orçamento dos serviços hospitalares.

No manual técnico do SIH (BRASIL, 2017), as AIH são classificadas em dois tipos:

- AIH 1 (Inicial) - para internamento inicial;
- AIH 5 (de Continuidade) - para casos de longa permanência, como transtornos psiquiátricos.

¹ Disseminação de Informações do Sistema de Informações Hospitalares (SIH) - Informe Técnico referente ao processamento 2016-03. Disponível em: ftp://ftp.datasus.gov.br/dissemin/publicos/SIHSUS/200801_Doc/IT_SIHSUS_1603.pdf

² Disponível em: <http://datasus.saude.gov.br/informacoes-de-saude>

AIH inicial é decorrente da emissão de laudo que após preenchimento é encaminhado ao autorizador, que pode exigir dados adicionais ou autorizar a internação de imediato. Com isso, o órgão emissor fornece o número da AIH juntamente com a identificação do autorizador.

AIH de continuidade admite um longo período de tratamento, geralmente psiquiátricos, tratamento de tuberculose entre outros. Neste caso, a autorização é “renovada” mês a mês durante a execução do tratamento.

2.2.1 Descrição da AIH

A AIH inicial é definida como o registro para os pacientes que precisam de tratamento em regime de internação. Seguindo as seguintes subclassificações: Principal, Especial e Secundária (BRASIL, 2017).

- Principal – Permite o registro de procedimentos que precisam de autorização e geram registro de AIH;
- Especial – Permite o registro do procedimento que necessita de autorização, porém não geram registro de AIH;
- Secundário – Permite o registro de procedimentos que não precisam de autorização e não geram registro AIH.

De acordo com o tipo de AIH que foi gerada, as informações são repassadas para devida aprovação. Conforme as notas técnicas divulgadas no portal do DATASUS (BRASIL, 2019c), as unidades hospitalares participantes do SUS (públicas ou particulares em convênio) enviam mensalmente os dados referentes às internações efetuadas através da AIH para os gestores municipais ou estaduais. Logo após, estas informações são processadas no DATASUS e são integradas ao Banco Nacional de Dados da saúde.

O arquivo DBC disponibilizado pelo DATASUS é estruturado em 113 variáveis preenchidas com dados dos tipos caractere e numérico. Ressaltando algumas dessas variáveis, suas definições são apresentadas de acordo com a documentação do SIH/2016-03:

- Ano e mês de processamento
 - Período do processamento da AIH, corresponde geralmente ao mês da alta.
- Local de Internação
 - Local (município, região metropolitana, microrregião, aglomerado urbano, regional de saúde, macrorregional de saúde, UF ou região) onde a unidade hospitalar está estabelecida.

- Município de Residência do Paciente
- Data de nascimento do paciente
- Sexo do paciente
- Dias de Permanência
 - Total de dias referentes ao período de internação. Exemplo: internação = 15/02/2019; alta = 18/02/2019; dias de permanência = 3.
- Diagnóstico principal
 - Representado pelo código do diagnóstico principal.

Em relação ao código do diagnóstico principal, este é determinado pela Classificação Internacional de Doenças e Problemas Relacionados com a Saúde (CID), estabelecida pela Organização Mundial de Saúde (OMS). A referente classificação fornece códigos relativos às doenças e uma grande variedade de sinais, sintomas, aspectos anormais, queixas, circunstâncias sociais e causas externas para ferimentos ou doenças (OMS, 2015). CID-10 corresponde a 10ª atualização dessa classificação.

O diagnóstico da dengue é identificado na CID-10 pelos seguintes códigos:

- **A90** – Dengue clássica;
- **A91** – Febre hemorrágica devida ao vírus da dengue.

A definição da CID-10 permite que os SIH possam referenciar as doenças de forma padronizada, facilitando a busca das informações diagnósticas nas bases de dados.

2.3 Ciência de Dados

O conceito de dados, conforme Medri (2011), pode ser compreendido como as informações inerentes às variáveis que caracterizam os elementos de uma população ou a amostra em análise. A aplicação da estatística sobre determinado conjunto de dados permite a descoberta de padrões, entre inúmeros outros resultados que podem auxiliar no planejamento estratégico de diversas áreas.

A evolução dos sistemas computacionais impulsionou o registro de dados correspondentes a inúmeros contextos de aplicação. As operações de controle, cadastro e automação de equipamentos em redes ocasionou um aumento significativo na dimensão das bases de dados (TRONCHONI et al., 2010). Segundo Porto e Ziviani (2014), o desafio emergente sobre a manipulação dos grandes volumes de dados corresponde a extração generalizada das informações que sejam relevantes para a definição de padrões que possam

auxiliar na tomada de decisão. A partir da manipulação dos dados com o objetivo de extrair conhecimento emergiu uma nova ciência, denominada Ciência de Dados.

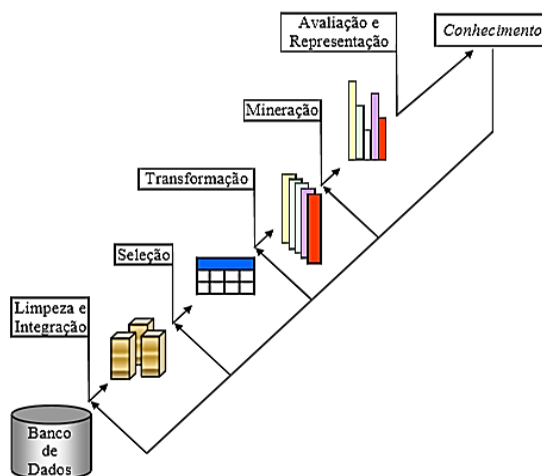
A Ciência de Dados consiste de vários elementos e aplicação de técnicas e teorias derivadas de múltiplos campos em engenharia e ciências. Objetivando-se a identificação dos princípios, métodos e técnicas fundamentais para o gerenciamento e análise de grandes volumes de dados. Dessa forma, os estudos correspondentes à Ciência de Dados tornaram-se importantes nas diversas áreas de conhecimento, tais como saúde, aplicações financeiras, astronomia, bioinformática, mobilidade urbana, segurança digital, comunicação e entre outras. (PORTO; ZIVIANI, 2014)

O processo de extração de informações úteis em enormes quantias de dados é definido através de técnicas e ferramentas que proporcionam a descoberta de conhecimento acerca de um determinado contexto. De acordo com Fayyad, Piatetsky-Shapiro e Smyth (1996), este processo é definido como Descoberta do Conhecimento (Em inglês, *knowledge-discovery in databases* - KDD), no qual aplicam-se os métodos de mineração de dados para extrair conhecimento e descobrir padrões.

Considerando que a informação é o ativo mais importante de uma organização, o processo de KDD possibilita que as decisões sejam tomadas visando estratégias com minimização de erros (DANTAS et al., 2008).

Na Figura 3 é possível visualizar a identificação das etapas do processo de KDD, conforme Han e Kamber (2000).

Figura 3 – Etapas do processo de KDD



Fonte: Tronchoni et al. (2010)

As etapas representadas na Figura 3 são descritas nos seguintes itens:

1. Limpeza dos dados: Consiste na remoção de ruídos e dados irrelevantes;
2. Integração dos dados: Outras fontes de dados múltiplos podem ser combinadas;
3. Seleção dos dados: Etapa na qual os dados relevantes para a análise são recuperados do banco de dados;
4. Transformação dos dados: Os dados são transformados ou consolidados no formato apropriado para mineração;
5. Mineração de dados: É um processo no qual se utiliza métodos inteligentes para extrair padrões de dados;
6. Avaliação e representação do conhecimento: As técnicas de visualização e representação de conhecimento são utilizadas para apresentar o conhecimento extraído.

Em virtude dos objetivos definidos para a extração do conhecimento, é muito importante que o domínio dos dados seja compreendido pelos profissionais da análise, com o propósito de selecionar os subconjuntos corretos de dados. Porém se ressalta que os problemas básicos de inferência estatística e descoberta de informações na manipulação de grandes volumes de dados são difíceis e desafiadores em suas mais variadas aplicações (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Desta forma, o uso de máquinas que automatizam a construção de modelos analíticos, identificando padrões, com o mínimo de intervenção humana, constituindo assim o objetivo principal da aprendizagem de máquina (SAS, 2019).

Conforme Han e Kamber (2000) explicitam, a estatística associada com as técnicas de visualização de dados possibilitam a descoberta do conhecimento por meio da exploração de dados. De acordo com a natureza do problema em análise, a aplicação das técnicas citadas constitui uma base para o pré-processamento e o Aprendizado de Máquina.

2.3.1 Análise Exploratória de Dados

A Análise Exploratória de Dados, antes definida como Estatística Descritiva, é constituída de métodos que permitem Refinar e Organizar os dados com o objetivo de encontrar regularidade ou padrão (MEDRI, 2011).

Segundo Medri (2011), a Estatística que também é compreendida como a ciência dos dados, se divide em duas áreas: Descritiva e a Inferencial (Indutiva). Para outros autores, essa ciência pode ser definida em três áreas: Descritiva, Probabilidade e Inferência estatística. De

todo modo, a Estatística consiste em um conjunto de técnicas para planejar experimentos, obter dados, organizá-los, resumi-los, analisá-los e interpretá-los.

Destacando a Estatística Descritiva, essa área é compreendida pelas etapas de organização, apresentação e sintetização de dados. Para isso, o uso de gráficos, tabelas e medidas descritivas permitem, na etapa inicial da análise, que as informações sejam apresentadas e posteriormente utilizadas em outras fases (MEDRI, 2011).

O processo de resumo e organização dos dados é constituído basicamente da ação de registrar as ocorrências dos valores correspondentes às variáveis que caracterizam um fenômeno. A distribuição de frequências permite que os dados sejam discriminados e agrupados conforme requisitos definidos no estudo (REIS, 2008). Quando se trata de grandes arquivos de dados, tal processo só é viável com o uso de processamento computacional.

Diante dos objetivos de uma análise de dados é importante ressaltar que a escolha da melhor técnica deve ser definida de acordo com o escopo da pesquisa. Alguns fatores como o número de variáveis, nível de mensuração das variáveis, tamanho do conjunto de dados, tempo disponível para apresentação dos resultados e público alvo influenciam no desenvolvimento do trabalho e na definição dos resultados.

2.4 Série Temporal

Durante o processo de análise de um conjunto de dados é possível relacionar a distribuição das variáveis conforme alguns parâmetros relevantes para o estudo. Um parâmetro relevante em diversas pesquisas quantitativas é o tempo. Verificar o comportamento de uma variável em um período de tempo pode ser definida como a análise de uma Série Temporal (ANTUNES; CARDOSO, 2015). Uma série temporal é conceituada como “sequências de dados quantitativos relativos a momentos específicos e estudados segundo sua distribuição no tempo” (WIENER, 1966)

A análise de séries temporais em diversas áreas do conhecimento é capaz de auxiliar a tomada de decisão para diferentes finalidades. Alguns exemplos de aplicações podem ser identificados no setor da Economia (Preços diários de ações, taxa de desemprego), Medicina (Níveis de eletrocardiograma), Epidemiologia (Casos semanais de sarampo e dengue) e Meteorologia (temperatura diária; registro de mares).

As séries temporais possuem três padrões básicos: Tendência, Sazonalidade e Ciclo. A tendência é definida de acordo com a maneira como os dados crescem ou diminuem ao longo

do tempo. Levando em consideração também que algumas séries podem não apresentar tendência, ou seja, os dados permanecem constantes ao longo do tempo (MARTIN et al., 2015).

A sazonalidade corresponde a frequência que determinado fenômeno se repete na série, podendo ser observado em um período sempre conhecido. Diferentemente das variações cíclicas que ocorrem em períodos não conhecidos. Enquanto as variações sazonais se completam no período de um ano, as variações cíclicas ocorrem em intervalo superior a um ano (MARTIN et al., 2015).

Com relação ao conjunto de dados ordenados em uma linha temporal, ele pode ser definido como discreto - por exemplo: o número mensal de casos notificados de uma doença específica - ou contínuo, como o registro dos valores de temperatura e umidade ao longo do dia (LATORRE; CARDOSO, 2001).

A partir da manipulação de um determinado conjunto de dados, visando a modelagem de um fenômeno, o uso das técnicas estatísticas possibilita que o comportamento de uma série temporal seja descrito e avaliado, assim como os fatores que a influenciam. Nos modelos de regressão polinomial, os valores de uma série são considerados como variável dependente (Y), enquanto os períodos em observação constituem a variável independente (X) (LATORRE; CARDOSO, 2001).

Manipular variáveis correspondentes a uma série temporal requer o auxílio de ferramentas de software estatístico que permitem construir graficamente a série de dados e são capazes de ajustar os valores para serem trabalhados em modelos preditivos. Alguns pacotes estão disponíveis nas linguagens Python e R, e podem ser utilizados gratuitamente em análises de dados.

2.5 Linguagem R

A exploração de dados corresponde a uma série de processos nos dados com o objetivo de gerar hipóteses, testá-las diversas vezes (WICKHAM; GROLEMUND, 2017). Neste contexto, os recursos computacionais possibilitaram a automatização das análises, e a manipulação dos dados com o auxílio de linguagens de programação, como o R e Python permitiu que os resultados fossem definidos com rapidez e precisão.

Ressaltando-se o R, é uma linguagem de programação interpretada, orientada a objeto e um ambiente computacional. Desenvolvida por Ross Ihaka e Robert Gentleman, na Universidade de Auckland, Nova Zelândia, a linguagem recebeu o nome de R por causa das iniciais de seus criadores. Além disso, o R é baseado nas linguagens S e LISP. Destaca-se ainda

que desde 1997, o núcleo internacional do R transformou a linguagem em um software de estatística e análise de dados sofisticado e completo (GOUVEIA, 2017).

O R é *open source*³ e suas bibliotecas recebem aprimoramento contínuo por parte da comunidade de desenvolvedores. Além disso, a linguagem permite a integração com códigos em C, C++, FORTRAN e Python. As funcionalidades disponíveis no R permitem a construção de gráficos com qualidade elevada, além de quase 10 mil pacotes com bibliotecas e métodos implementados. (CHIAVEGATTO FILHO, 2015; GOUVEIA, 2017).

Obedecendo aos conceitos de orientação a objeto, no R é possível criar Classes e instanciar objetos. No qual, um objeto pode armazenar dados e acessar os métodos implementados em uma classe (DEITEL; DEITEL, 2016). Os tipos básicos de estruturas no R são:

- **Atomic vector:** Estruturas homogêneas e unidimensionais, definidas como lógico, inteiro, flutuante, complexo e caractere;
- **Array:** Estrutura homogênea e unidimensional;
- **Matriz:** Estrutura homogênea e bidimensional;
- **Lista:** Estrutura heterogênea, função de chamada: `list()`;
- **Dataframe:** Estrutura heterogênea, na qual cada coluna representa uma variável e cada linha corresponde a uma observação. São definidos usando a função `data.frame()`.

Para a análise estatística, a linguagem R fornece funções e métodos eficazes. Como exemplo, destaca-se a função *summary*, que permite retornar o resumo de cada variável quando aplicada a um banco de dados. Nesta função ocorre o agrupamento de outras funções como Valor Mínimo, Valor Máximo, Média, Mediana e Quartis.

Através de outras funções estatísticas é possível calcular no R medidas de dispersão amostral (Variância, Desvio padrão e Amplitude total), covariância, correlação e distribuição de probabilidade.

Para a representação dos dados, é possível construir com o R uma variedade de gráficos e figuras. A função `plot()` permite formatar o desenho do gráfico de acordo com o ajuste dos parâmetros. O principal recurso para controlar o aspecto dos gráficos é a função de configuração

³ *Open source* é uma licença de código aberto, no qual a autoria fornece o direito de estudar, modificar e distribuir o software de graça para qualquer um e para qualquer finalidade (RED HAT Inc, 2019).

par()). Com esta função pode-se configurar o formato, tamanho, subdivisões, margens, entre diversas outras opções.

Para produzir gráficos em arquivos basta redirecionar o dispositivo de saída para o formato desejado, utilizando funções como `pdf()` ou `jpeg()`. Cada uma dessas funções possui argumentos adicionais para controlar tamanho, resolução, entre outros atributos do arquivo. Além disso, com a função `axis()` é possível redesenhar alguma parte do gráfico, caso seja necessário.

A linguagem R dispõe de pacotes que permitem a leitura de arquivos em bases de dados. Com o pacote *readr* é possível importar arquivos do tipo ‘CSV’, ‘TSV’ e ‘FWF’. Se o arquivo estiver estruturado no formato de planilha eletrônica (XLS ou XLSX), é possível utilizar o pacote *readxl* para realizar a importação dos dados. Uma das maiores vantagens desses pacotes para a manipulação dos dados é que bases enormes podem ser acessadas diretamente do disco rígido sem prejudicar o processamento da máquina (ITANO; SANTOS, 2007).

O R também dispõe de um pacote que permite a leitura de bases de dados no formato DBC, trata-se do pacote *read.dbc*. Os arquivos com este formato são disponibilizados pelo DATASUS contendo dados nacionais sobre alguns índices de saúde pública, como é o caso das internações hospitalares.

A extensão de arquivo DBC foi desenvolvida pelo DATASUS para facilitar o armazenamento das enormes quantias de dados que eram produzidos pelos diversos sistemas de saúde pública. Porém, estes arquivos só podiam ser lidos através de ferramentas disponibilizadas pelo Ministério da Saúde. Com isso, o pacote *read.dbc* foi desenvolvido como uma extensão da linguagem R para facilitar a leitura e a manipulação dos dados compactados neste formato (PETRUZALEK, 2016).

2.6 Trabalhos Relacionados

Wanderley (2018), apresenta em seu trabalho a análise dos dados da dengue por meio de séries temporais acerca da relação entre a doença, fatores socioeconômicos, ambientais e as variáveis climáticas no município de Patos-PB. Para isso, foram utilizados os dados referentes aos casos de dengue e do clima disponibilizados pelas bases do Sistema de Informação de Agravos de Notificação (SINAN) e do Instituto Nacional de Meteorologia (INMET). Identificou-se a evidência dos fatores relacionados a estratificação dos níveis sociais e das variáveis ambientais, principalmente o saneamento básico como preponderantes para a ocorrência dos casos de dengue. Além disso, a caracterização das variáveis climáticas

apresentou correlação entre a umidade relativa e precipitação com as alterações climáticas contínuas e a ocorrência de dengue. Ressalta-se também que no estudo, apesar de não ter sido possível indicar uma temperatura melhor relacionada com a ocorrência dos casos, foi possível verificar que os períodos mais quentes no município de Patos/PB apresentam maior correlação com o período sazonal da dengue.

Mendes Neto (2019) desenvolveu uma análise exploratória dos dados referentes à dengue no município de Patos, a fim de verificar relações entre fatores climáticos e os casos notificados de dengue, utilizando os dados disponibilizados pelo SINAN, INMET e da Agência Executiva de Gestão das Águas (AESAPB). Por meio da realização de análises utilizando as bibliotecas da linguagem de programação Python, e com auxílio de algoritmo de regressão linear e predição de dados verificou-se que os casos de dengue estão relacionados diretamente com as mudanças climáticas na cidade de Patos. Percebeu-se que a maior incidência dos casos ocorreu no período entre março e julho, devido a diminuição da temperatura, aumento nos níveis de chuva e umidade do ar.

Nascimento et al. (2015) explana em sua pesquisa a descrição do perfil clínico-epidemiológico e a distribuição espacial dos casos suspeitos de dengue com hospitalização na cidade de Goiânia-GO durante a epidemia de 2013. Desenvolveu-se um estudo descritivo sobre dados do Sistema de Informação de Agravos de Notificação, Sistema Gerenciador de Ambiente Laboratorial, Sistema de Regulação de Leitos, e do Sistema da Companhia de Processamento de Dados de Goiânia, obtendo-se como resultados a identificação de 616 casos suspeitos de dengue com idade média de 36 anos, sendo 55,2% do sexo feminino. Também foi possível identificar mediante às análises realizadas a predominância de adultos com presença de sinais e sintomas de alarme da dengue.

Os trabalhos citados acima permitem o reconhecimento do cenário da dengue em duas regiões distintas do país. Com destaque para as pesquisas de Wanderley (2018) e Mendes Neto (2019), que concentraram as análises da doença na cidade de Patos, que é um dos 223 municípios da Paraíba. Os resultados desses estudos são importantes para verificar se algum dos fatores de agravo para a ocorrência da dengue também serão identificados nas áreas com maior concentração de internação hospitalar no Estado. Além disso, o trabalho de Nascimento et al. (2015) servirá como referência para se estruturar um possível perfil dos pacientes internados durante o período de janeiro de 2014 a dezembro de 2018.

3 METODOLOGIA

Este trabalho consiste de uma pesquisa aplicada com abordagem quantitativa, estruturada com objetivos descritivos e explicativos.

Segundo Fonseca (2002, p. 20):

A pesquisa quantitativa se centra na objetividade. Influenciada pelo positivismo, considera que a realidade só pode ser compreendida com base na análise de dados brutos, recolhidos com o auxílio de instrumentos padronizados e neutros.

Gerhardt e Silveira (2009) expõem que a pesquisa quantitativa enfatiza o raciocínio dedutivo e os atributos mensuráveis da experiência humana. E no que se refere a natureza de pesquisa aplicada, busca-se gerar conhecimento para aplicação prática objetivando solucionar problemas específicos.

Diante dos objetivos descritivos e explicativos, Gil (2008) aborda que uma pesquisa explicativa tem viés descritivo, em virtude que a identificação dos fatores influentes a um fenômeno impera descrição e detalhamento suficiente dos fatos.

Em relação ao levantamento bibliográfico, buscou-se identificar e interpretar estudos desenvolvidos com semelhante problemática em referência a dengue. E dessa forma, reconhecer as técnicas e metodologias aplicadas para a manipulação dos dados da doença.

Quanto a amostragem, os dados em análise pertencem às AIH do Estado da Paraíba do período de janeiro de 2014 a dezembro de 2018. Estes dados foram obtidos através do portal do DATASUS e inicialmente foram verificados com auxílio da ferramenta RStudio, utilizando a linguagem R para reconhecer as variáveis definidas e a completude das bases de dados.

A respeito da análise realizada buscou-se além de mapear os casos de internações, entender e caracterizar o comportamento dos casos ao longo dos meses com o auxílio de séries temporais e ainda se investigou a relação entre o número de internações e outras variáveis, como a densidade demográfica, o volume de chuva e o número de focos do mosquito vetor.

A densidade demográfica foi selecionada para análise, pois buscou-se entender como essa variável pode influenciar no número de casos de dengue de um município. Levando em consideração que se muitas pessoas vivem em uma mesma área, existe a possibilidade do total de infectados ser maior quando ocorre um surto.

A variável correspondente ao volume de chuvas foi considerada por estar entre uma das variáveis mais citadas nos estudos envolvendo fatores climáticos e a dengue.

E por último pretendeu-se verificar se nas áreas mais afetadas, também foi identificado condições de risco para a população por causa número de focos do mosquito *Aedes aegypti*.

A pesquisa ainda foi modelada conforme o *Design Science* (LACERDA et al., 2013), a partir da necessidade de responder às seguintes questões principais (QP) de pesquisa:

QP1: Quais foram as regiões da Paraíba mais afetadas pela doença dengue no período em análise?

QP2: Quais os períodos com maior número de internações hospitalares no Estado?

Relacionadas às questões principais de pesquisa apresentadas anteriormente, seguem as seguintes questões secundárias de pesquisa (QSP):

QSP1: Quais foram os municípios com os maiores registros de internações por dengue?

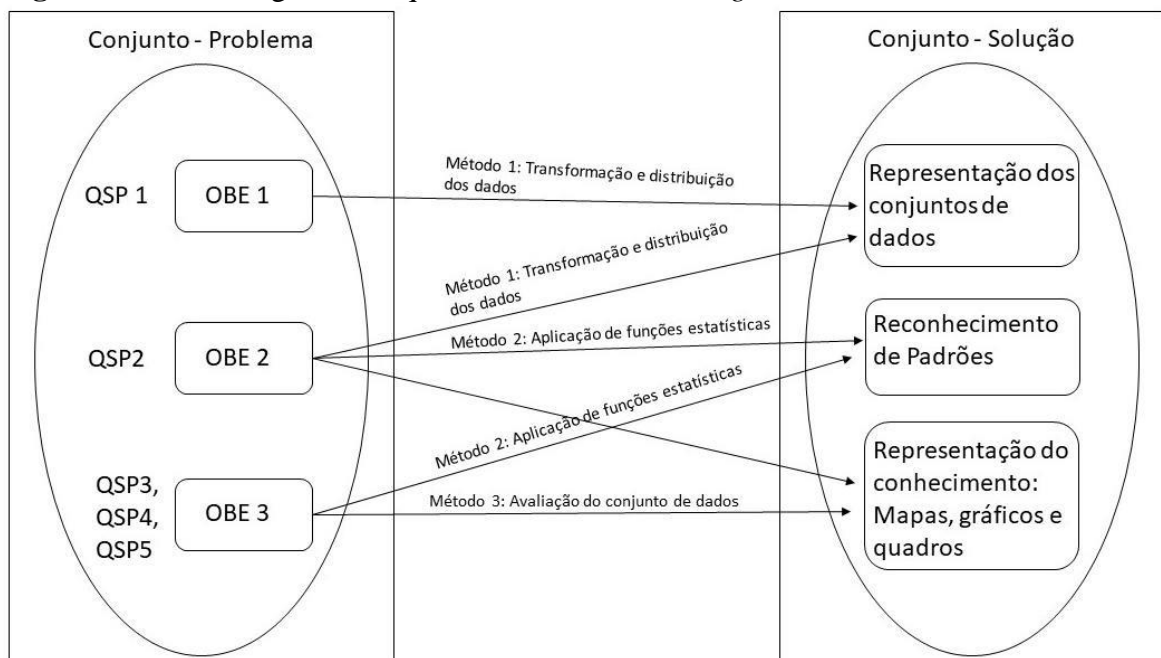
QSP2: Qual a sazonalidade dos casos no período observado?

QSP3: Qual a relação entre o número de internações e a densidade demográfica dos municípios mais afetados?

QSP4: Qual a relação entre o número de internações e o volume de chuvas das áreas mais vulneráveis?

QSP5: Qual a relação entre o número de internações e número de focos do mosquito vetor identificados nas regiões?

Figura 4 - Metodologia de Pesquisa de acordo com *Design Science*



Fonte: A autora (2019).

Os métodos destacados na Figura 4 correspondem ao processo adotado para obter o conjunto solução, conforme os objetivos da pesquisa. A seguir é possível verificar a especificação de cada um deles no processo exploratório.

- Método 1: Transformação e distribuição dos dados – Os dados foram refinados e organizados para formar conjuntos de valores específicos conforme restrições definidas.
- Método 2: Aplicação de funções estatísticas – Os conjuntos de dados foram submetidos ao processamento através de funções estatísticas, para se reconhecer os padrões quantitativos.
- Método 3: Avaliação do conjunto de dados – Os dados foram observados, mensurados e representados conforme os objetivos definidos. Possibilitando a compreensão dos resultados obtidos com a análise.

3.1 Coleta de dados

O principal foco desta pesquisa foi analisar os dados referentes às internações hospitalares, com diagnóstico de dengue registradas no Estado da Paraíba. Logo, foi necessário criar um *dataset* (uma coleção de dados) utilizando os dados extraídos dos formulários de AIH para que as etapas posteriores da análise fossem realizadas.

Os dados foram obtidos através dos arquivos de bases de dados em formato DBC, disponibilizados pelo DATASUS mensalmente no portal de acesso: datasus.saude.gov.br/informacoes-de-saude/. Os arquivos de dados foram selecionados para download de acordo com o período definido nos objetivos da pesquisa. Assim, foram baixados os dados correspondentes aos registros de internações hospitalares na Paraíba de janeiro de 2014 a dezembro de 2018.

Os arquivos de AIH reduzidas são constituídos de todas as internações hospitalares registradas em um determinado mês, em todas as unidades federativas do Brasil.

Para se realizar a leitura inicial dos dados, foi necessário concatenar os arquivos das internações divididos em meses em um único *dataset*. A compilação dos dados foi possível através da execução de um *script* escrito na linguagem R, que realizava a instalação dos pacotes "*read.dbc*", "*stringr*" e "*lubridate*", além da leitura dos registros por meio de métodos e laços de repetição, adicionando-os em uma variável temporária que posteriormente foi convertida para um arquivo CSV e uma planilha eletrônica.

A respeito de outros dados que foram explorados, buscando-se identificar uma possível relação as internações, a densidade demográfica dos municípios e o volume de chuva foi necessário fazer consultas à outras bases específicas. Os dados da densidade demográfica dos municípios da Paraíba foram coletados do Instituto Brasileiro de Geografia e Estatística (IBGE)

atualizados conforme o censo demográfico do ano de 2010. Enquanto os dados referentes ao volume de chuva de cada município por ano de seleção foram extraídos do portal da AESA/PB.

3.1.1 Tratamento dos dados

Os dados foram tratados com o auxílio de dois softwares, o RStudio, que é um ambiente de desenvolvimento integrado para o R e o Excel 2016, que se trata de software editor de planilhas eletrônicas. O mesmo conjunto de dados foi trabalhado em ambos os softwares, porém com objetivos distintos. A manipulação dos dados com RStudio foi necessária para formatar o conjunto e construir os gráficos. Enquanto o tratamento dos dados no Excel foi essencial para o desenvolvimento dos mapas graduados.

Depois que o script foi executado, o *dataset* ficou composto por 113 variáveis da AIH e 6926 linhas de registros. Porém, antes do arquivo CSV e da planilha serem construídos, o total de variáveis foi reduzido para apenas 14 variáveis, as demais foram excluídas por não apresentarem dados relevantes para a pesquisa.

Os dados armazenados no arquivo CSV foram importados no RStudio, por meio do método *read.csv* do R. Com a importação dados em um novo *dataset*, foi possível manipular as variáveis e assim identificar os padrões correspondentes aos casos de dengue e representá-los por ano e mês de ocorrência em gráficos de linha.

Com relação ao processo desenvolvido no Excel, o arquivo de planilha eletrônica original foi subdividido em cinco outras planilhas, cada uma corresponde a um ano de ocorrências, no caso 2014, 2015, 2016, 2017 e 2018. Este processo foi fundamental para formatar os dados e aplicar as funções estatísticas, além de possibilitar o ajuste dos dados para serem trabalhados no QGIS 3.4, que é um software para processamento de informação geográfica.

É importante ressaltar que no processo de construção dos mapas no software QGIS foi necessário realizar um novo ajuste no conjunto de dados. Pois o *shapefile*⁴ dos municípios da Paraíba disponibilizado pelo IBGE processa o código dos municípios como inteiros de sete dígitos, enquanto na planilha das internações hospitalares, o código de cada município estava preenchido como um inteiro de seis dígitos. O ajuste nos códigos permitiu que a função de união do QGIS gerasse uma tabela de atributos essencial para a graduação dos mapas.

⁴ *Shapefile* é um formato de arquivo contendo dados geoespaciais em forma de vetor usado por Sistemas de Informações Geográficas (SIG) (ESRI, 1998).

3.2 Tabulação dos dados

Após a seleção e tratamento dos dados, a planilha eletrônica original e o arquivo CSV ficaram estruturados em 14 colunas e 6.297 linhas. No Quadro 1 é possível identificar a especificação dos tipos de dados referentes às variáveis utilizadas.

Quadro 1 – Tipos de variáveis do conjunto de dados

| Variável | Tipo de dado | Observação |
|-------------|---------------------|----------------------------------|
| MUNIC_RES | Inteiro (6 dígitos) | - |
| MUNIC_NOME | Char | - |
| IDADE | Inteiro | 0 anos – 99 anos |
| DIAG_PRINC | Char | A90 (CID da Dengue) |
| DIAS_PERM | Inteiro | Mínimo: 0 dia Máximo: 33 dias |
| SEXO | Char | F ou M |
| MES_INTER | Inteiro | 1 - 12 |
| ANO_INTER | Inteiro | 2014 - 2018 |
| COD_UF | Inteiro | 25 (IBGE) |
| UF_NOME | Char | Apenas Paraíba |
| DT_INTER_BR | Data (dd/mm/aaaa) | - |
| FAIXA_ETA | Char | - |
| LATITUDE | Inteiro | Correspondente ao município |
| LONGITUDE | Inteiro | Correspondente ao município |

Fonte: A autora (2019).

No capítulo 4 serão apresentados os resultados e as discussões pertinentes aos objetivos da pesquisa, com base na análise exploratória realizada.

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Neste capítulo serão discutidos os resultados da análise dos dados das internações hospitalares com diagnóstico de dengue, apresentando o mapeamento das áreas mais afetadas pela doença, assim como a relação entre dados observados por meio dos processos de descoberta de conhecimento em base de dados.

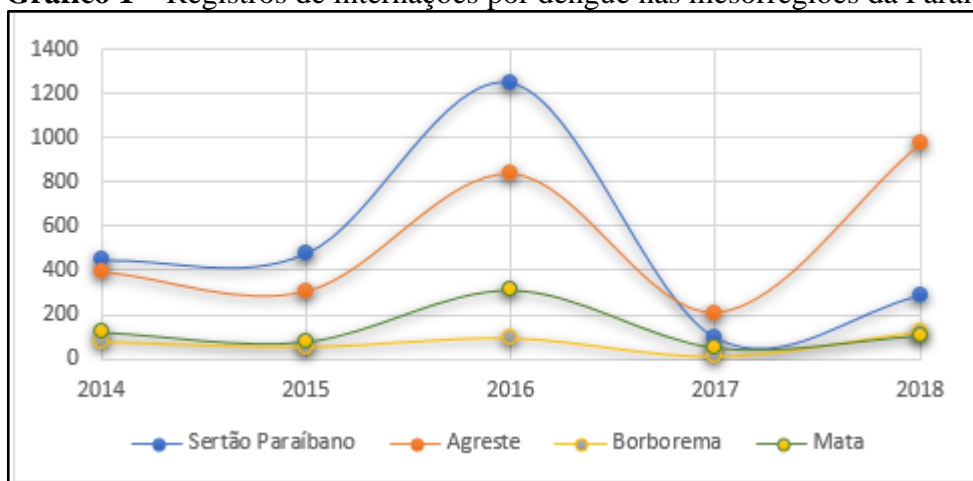
4.1 Casos de dengue no Estado da Paraíba

A análise dos dados de AIH permitiu identificar o total de internações por dengue em cada município da Paraíba, possibilitando o reconhecimento da distribuição dos casos que foram registradas no período de janeiro de 2014 a dezembro de 2018.

A partir do reconhecimento do número de casos foi possível agrupar os dados por mesorregião do Estado, as quais são denominadas como Sertão paraibano, Borborema, Agreste paraibano e Mata Paraibana, e dessa forma identificar que as mesorregiões mais afetadas pela dengue no período observado foram o Sertão e o Agreste Paraibano.

O Gráfico 1 apresenta o número de casos de internações com diagnóstico de dengue nas mesorregiões durante os cinco anos analisados.

Gráfico 1 – Registros de internações por dengue nas mesorregiões da Paraíba

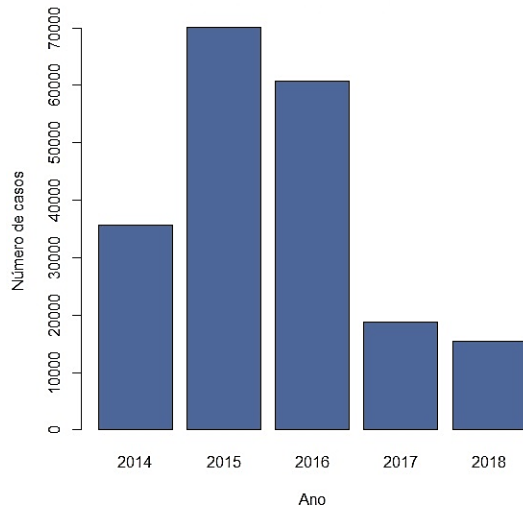


Fonte: A autora (2019)

No Gráfico 1 percebe-se que o ano de 2016 concentrou o maior número de internações por dengue no Estado, foram registrados 2487 casos. Também é possível ressaltar o aumento de internações no ano de 2018, quando comparado ao ano de 2017. Foram registradas 1494 internações em 2018, enquanto em 2017 registrou-se 365 internações. Isso indica que houve um crescimento de aproximadamente 309%.

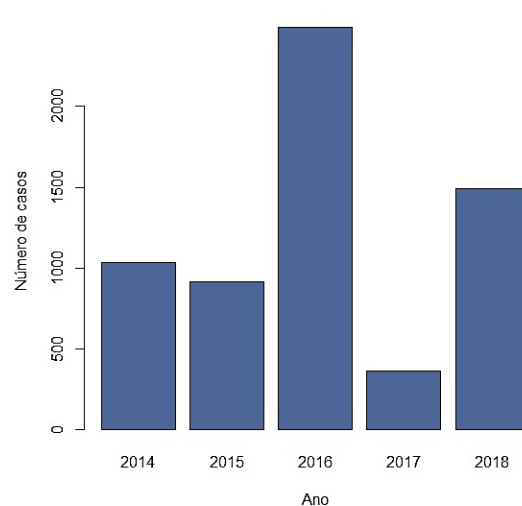
Ao avaliar o total de internações da Paraíba, buscou-se verificar se o padrão identificado no Estado era semelhante ao padrão de casos registrados no Brasil no mesmo período. Nos Gráficos 2 e 3 são apresentados o número de casos por ano no território nacional (todas as unidades federativas) e na Paraíba.

Gráfico 2 - Internações por dengue - BR



Fonte: A autora (2019)

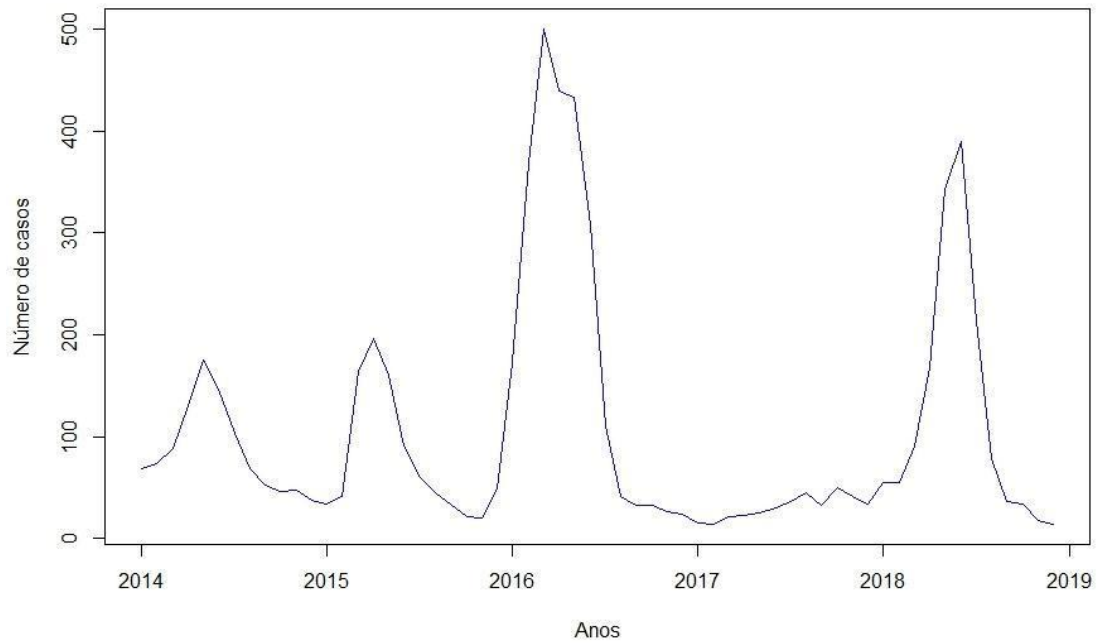
Gráfico 3 - Internações por dengue -PB



Fonte: A autora (2019)

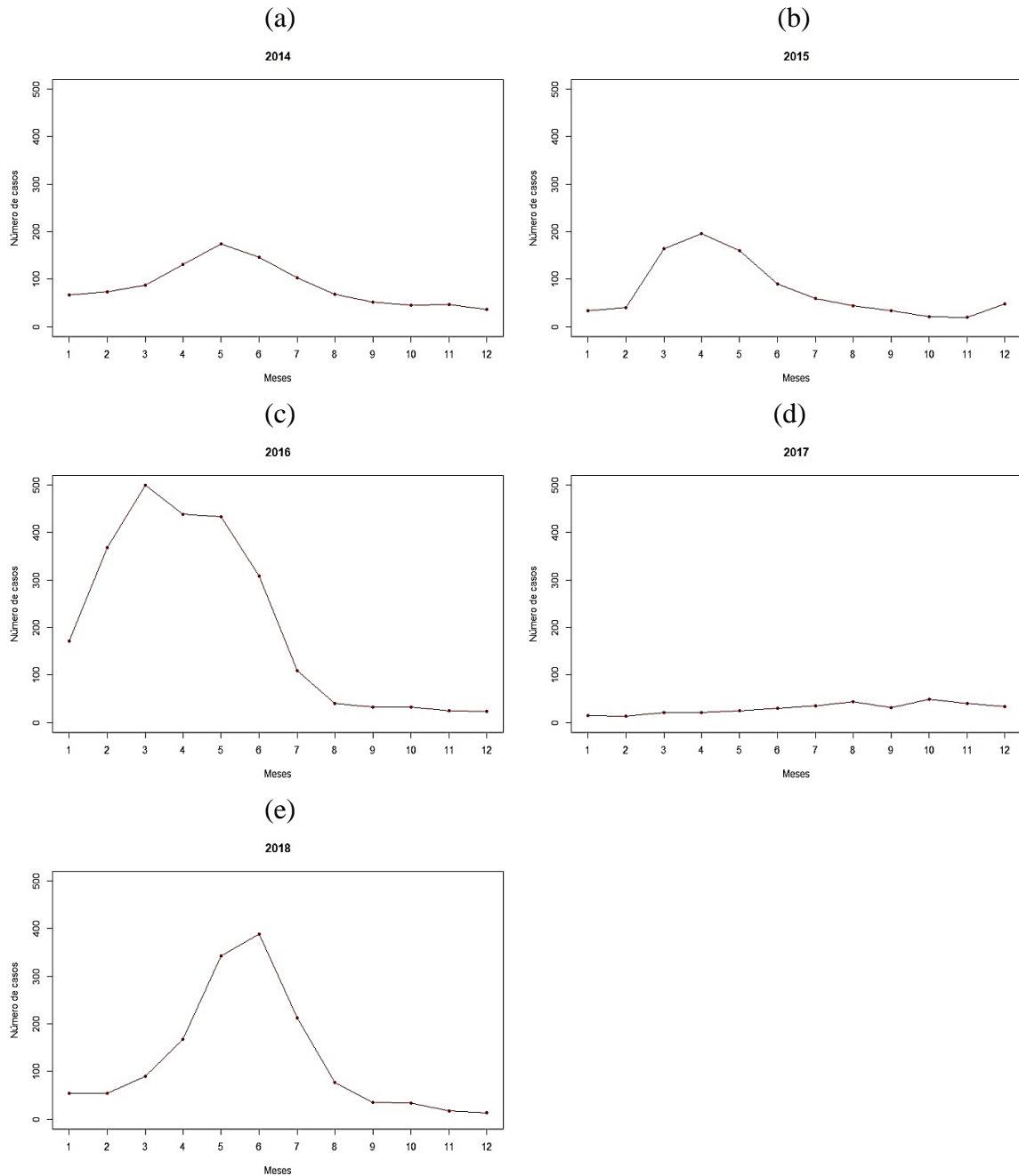
Observa-se que o padrão de crescimento no número de internações registradas na Paraíba não coincide com o comportamento dos casos registrados no Brasil. É possível verificar que a maior concentração de internações no Brasil ocorreu no ano de 2015, enquanto na Paraíba esse resultado foi identificado no ano de 2016. Além disso, em nível nacional, o número de internações por dengue reduziu muito nos anos de 2017 e 2018, porém em 2018 o total de casos na Paraíba voltou a subir, e este é o segundo período com mais registros de casos no Estado.

Considerando a data de registro em cada internação foi possível construir uma série temporal ordenada pelos 12 meses de cada ano avaliado, se ressaltando a tendência dos casos conforme a decorrência do tempo. No Gráfico 4, a série temporal apresentada evidencia o volume de internações, principalmente nos anos de 2016 e 2018.

Gráfico 4 – Casos de internações por ano, período de 2014 - 2018

Fonte: A autora (2019)

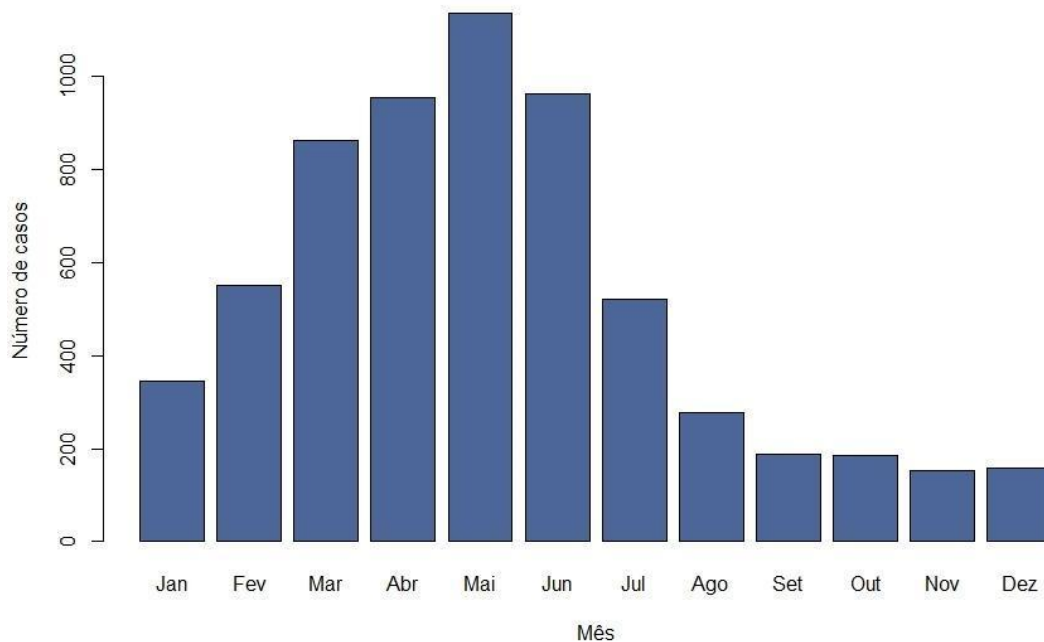
Diante do resultado observado no Gráfico 4, desenvolveu-se ainda a avaliação dos dados conforme a incidência de casos nos meses de cada ano, buscando-se identificar os períodos de maior risco de acordo os dados das internações. O Gráfico 5 apresenta o crescimento e a diminuição do número de casos conforme a progressão dos meses durante os cinco anos.

Gráfico 5 – Variação dos casos de internações em meses, período de 2014 – 2018

Fonte: A autora (2019)

Levando em consideração principalmente a variação dos Gráficos 5(a), 5(b), 5(c) e 5(e) observa-se que os registros de internações tendem a apresentar crescimento no primeiro semestre dos anos. Com exceção nos casos do Gráfico 5(d) no qual é possível identificar um baixo crescimento de casos nos últimos seis meses do ano de 2017.

No Gráfico 6 é possível notar uma concentração maior de casos durante os meses de março a junho. Ressaltando-se que a análise foi realizada em 60 meses, considerando a distribuição dos 12 meses em cada um dos cinco anos.

Gráfico 6 – Concentração dos casos de internações por meses, período de 2014 – 2018

Fonte: A autora (2019)

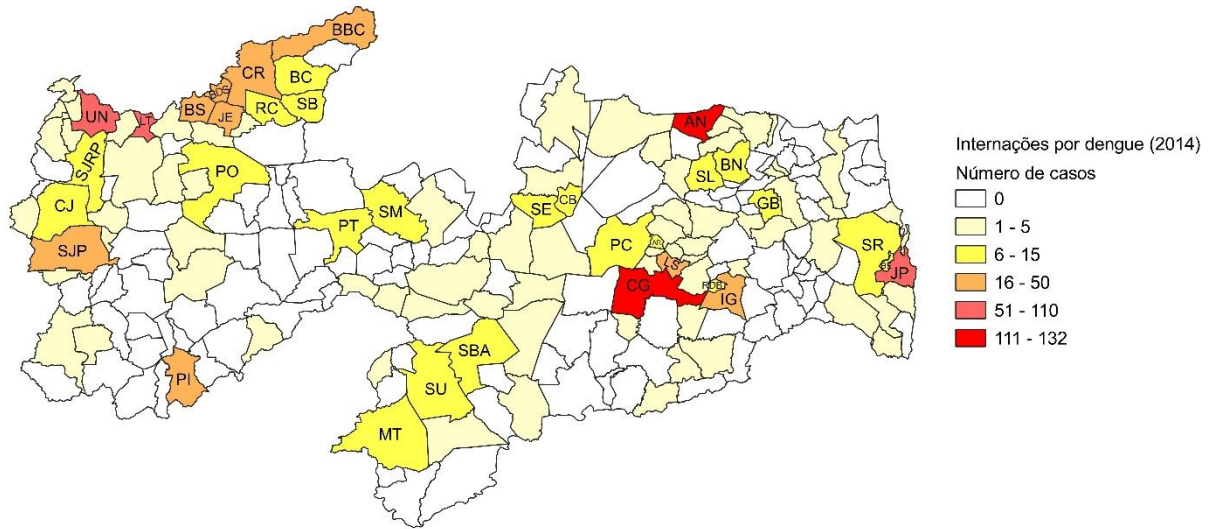
Percebe-se no Gráfico 6 que o número de internações tende a diminuir com o passar dos meses em um ano. É possível considerar que esse comportamento seja resultante da variação climática do Estado, pois diversos outros estudos (VIANA; IGNOTTI, 2013; VINCENTI-GONZALEZ et al., 2018) sobre a dengue apontam que a variação do clima em uma região é um dos fatores de maior influência para o aumento de casos da doença.

4.2 Mapeamento das áreas de risco no Estado da Paraíba

Através da análise realizada foi possível contabilizar as internações de acordo com o município de residência do paciente. Com a distribuição dos dados em mapas graduados identificou-se quais foram os municípios da Paraíba mais afetados pela doença dengue durante o período de 2014 a 2018.

O Mapa 1 representa a distribuição das internações no ano de 2014, e as cores ressaltadas nos municípios indicam a categoria de limite na contagem de casos registrados. Em relação às abreviaturas de identificação para o nome dos municípios (ver Lista de Abreviaturas), decidiu-se rotular apenas as cidades que apresentassem casos suficiente para serem destacadas entre a terceira e sexta categoria limite, de acordo com o ajuste do software.

Mapa 1 – Identificação dos casos de internações por município (2014)

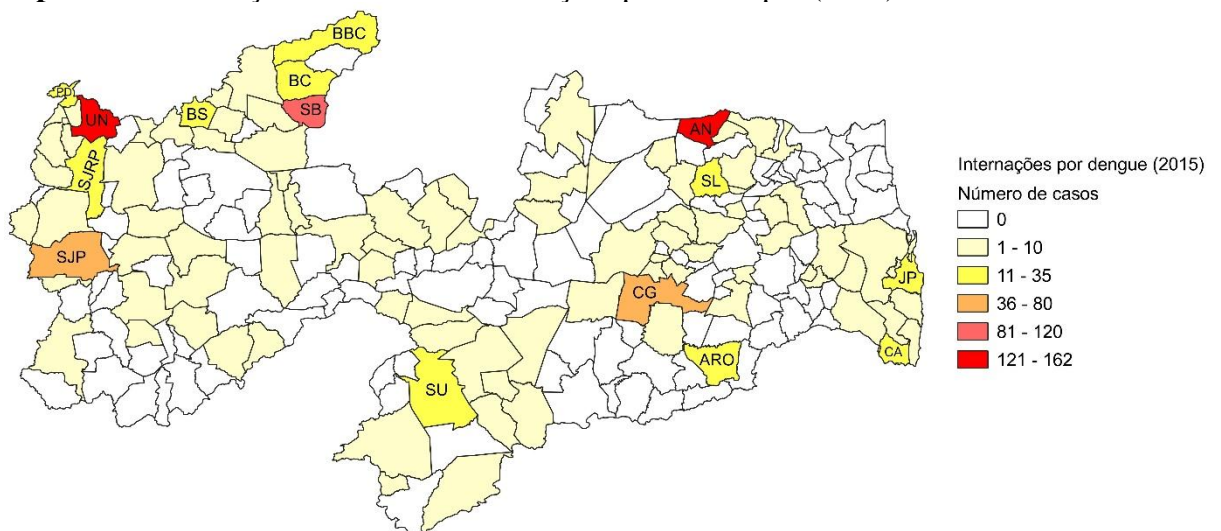


Fonte: A autora (2019)

No Mapa 1 é possível observar que no ano de 2014 os municípios mais afetados pela doença dengue foram os destacados com mais de 50 internações registradas. Sendo eles o município de Uiraúna (UN), Lastro (LT), Araruna (AN), Campina Grande (CG) e João Pessoa (JP).

No Mapa 2 destaca-se a distribuição de casos durante o ano de 2015, e percebe-se que o número de casos no município de Campina Grande foi reduzido, porém os municípios de Araruna e Uiraúna ainda apresentaram os maiores registros quando comparado a outros municípios.

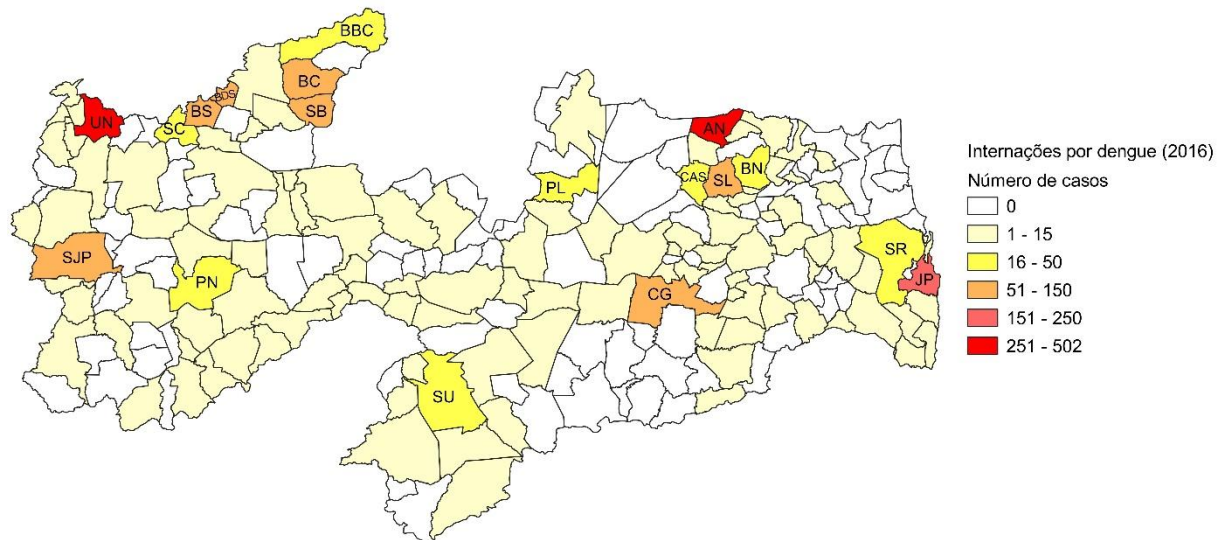
Mapa 2 – Identificação dos casos de internações por município (2015)



Fonte: A autora (2019)

O Mapa 3 foi construído conforme a distribuição das internações no ano de 2016. Como já foi apresentado anteriormente, no ano de 2016 foi registrado o maior número de internações no período da análise. Por isso, o limite das categorias de cor foi ampliado consideravelmente para se adequar aos valores verificados nas cidades.

Mapa 3 – Identificação dos casos de internações por município (2016)

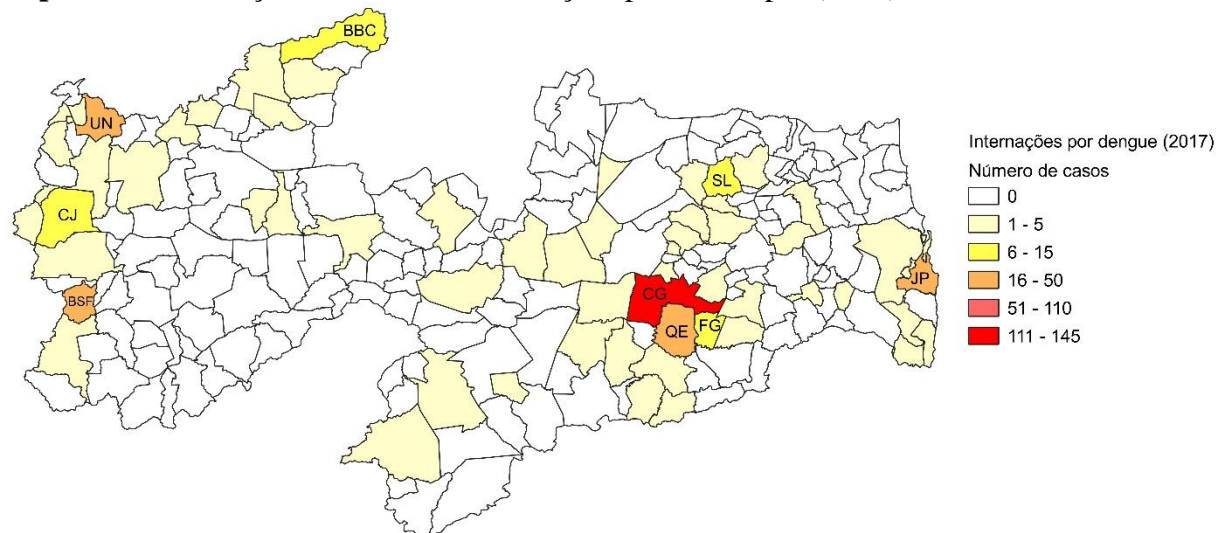


Fonte: A autora (2019)

Através do Mapa 3 apresentado, percebe-se que os municípios de Araruna e Uiraúna apresentaram as maiores taxas de internações por dengue no ano de 2016. Ressaltando-se ainda a elevação no número de casos no município de João Pessoa.

No Mapa 4 é possível notar que o total de internações no Estado diminuiu significativamente, considerando os dados referentes ao ano de 2016.

Mapa 4 – Identificação dos casos de internações por município (2017)

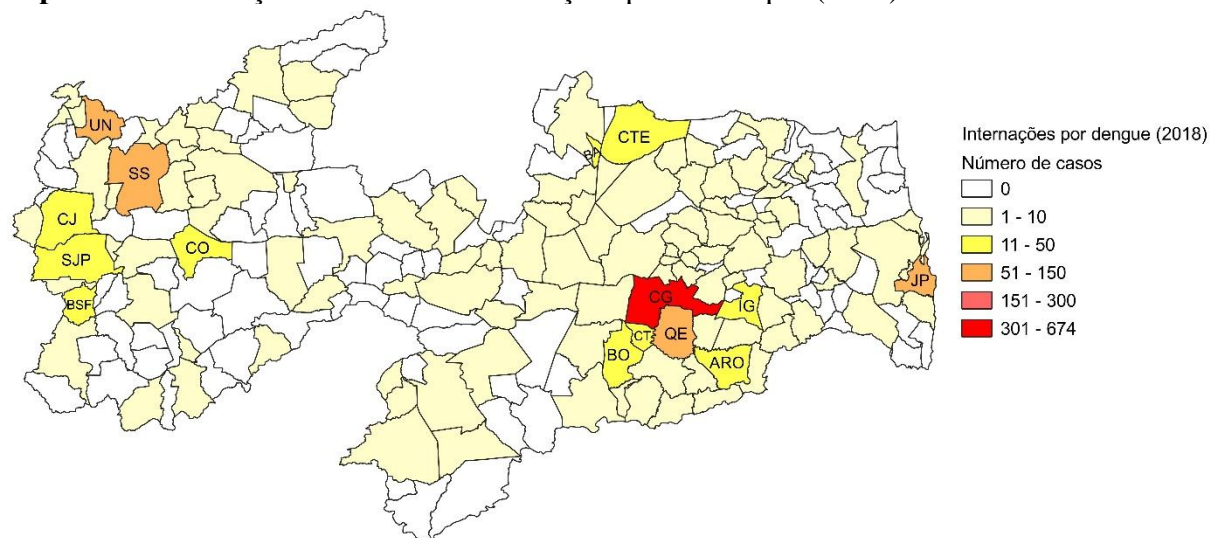


Fonte: A autora (2019)

No ano de 2017 foi possível verificar que as internações do município de Uiraúna diminuíram, em relação aos três anos anteriores. No total foram apenas 30 casos registrados. Além disso é preciso ressaltar que a cidade de Araruna não apresentou nenhuma subnotificação de internação por dengue. No entanto, a cidade de Campina Grande concentrou o maior valor de casos no ano.

No Mapa 5, referente à distribuição de internações em 2018, é possível observar um aumento de casos, em relação aos casos de 2017. E novamente a reincidência do maior valor de internações ser identificado no município de Campina Grande.

Mapa 5 – Identificação dos casos de internações por município (2018)



Fonte: A autora (2019)

Além do município de Campina Grande, é possível perceber ainda que as cidades de Uiraúna, João Pessoa, Queimadas (QE) e Sousa (SS) registraram mais de 50 internações no período, sendo identificado em Queimadas 121 casos, o segundo maior total de 2018.

Levando em consideração as informações obtidas por meio da análise realizada é possível afirmar que os municípios de Campina Grande, Araruna, Uiraúna e João Pessoa foram identificados como as áreas mais vulneráveis para a transmissão da dengue na Paraíba. Dessa forma, buscou identificar nestes municípios a relação entre o total de internações e as variáveis como densidade demográfica, volume de chuva e número de focos do mosquito observados na cidade.

4.3 Análise do número de internações em relação à densidade demográfica, volume de chuvas e aos focos do *Aedes aegypti*

Definiu-se o objetivo de identificar a relação entre o número de internações e as variáveis de densidade demográfica, volume de chuva e o número de focos do mosquito nos municípios. Para isso, foi selecionado em cada ano os cinco municípios com a maior contagem de internações, e estes foram avaliados conforme as variáveis apresentadas.

No Quadro 2 observa-se que os municípios de Araruna, Uiraúna, Campina Grande e João Pessoa se destacaram com os maiores registros de internações em pelo menos três dos cinco anos analisados.

Quadro 2: Municípios com maiores registros de internações (2014 – 2018)

| Município | Numero de Internações | Ano |
|----------------------|-----------------------|-----|
| Araruna | 132 | 2 |
| Campina Grande | 126 | 0 |
| Lastro | 96 | 1 |
| João Pessoa | 80 | 4 |
| Uiraúna | 71 | |
| Araruna | 162 | 2 |
| Uiraúna | 139 | 0 |
| São Bento | 86 | 1 |
| São José de Piranhas | 73 | 5 |
| Campina Grande | 65 | |
| Araruna | 502 | 2 |
| Uiraúna | 448 | 0 |
| João Pessoa | 238 | 1 |
| São Bento | 126 | 6 |
| Bom Sucesso | 115 | |
| Campina Grande | 145 | 2 |
| João Pessoa | 36 | 0 |
| Uiraúna | 30 | 1 |
| Bonito de Santa Fé | 24 | 7 |
| Queimadas | 23 | |
| Campina Grande | 674 | 2 |
| Queimadas | 121 | 0 |
| João Pessoa | 81 | 1 |
| Uiraúna | 74 | 8 |
| Sousa | 58 | |

Fonte: A autora (2019)

No Quadro 2 destaca-se principalmente a reincidência do município de Uiraúna de estar entre os cinco municípios mais afetados em todos os anos de 2014 a 2018.

Ao verificar uma possível relação entre a densidade demográfica e o número de internações dos municípios destacados, percebeu-se que os municípios de Araruna e Uiraúna são bem menos populosos que os municípios de João Pessoa e Campina Grande, e ainda assim eles foram bastante afetados pela doença dengue.

Na Tabela 1 é possível identificar os valores correspondentes aos municípios citados.

Tabela 1: Relação entre densidade demográfica e número de internações

| Município | Densidade (Hab/Km ²)* | Total de casos (2014 - 2018) |
|----------------|-----------------------------------|------------------------------|
| Campina Grande | 648,31 | 1100 |
| Araruna | 76,83 | 796 |
| Uiraúna | 49,52 | 762 |
| João Pessoa | 3421,28 | 466 |

*Dados extraídos do Censo 2010 – IBGE

Fonte: A autora (2019)

Antes dos resultados da análise foi considerado que os maiores registros de internações se concentrariam nas cidades de João Pessoa e Campina Grande, por causa que estas são as duas cidades mais populosas do Estado, o que poderia ter influenciado em um número maior de casos. No entanto, essa relação não foi definida devido ao total de casos registrados nos municípios de Araruna e Uiraúna. Pois, apesar da baixa densidade demográfica destas cidades, a contagem de internações ocorridas com seus residentes foi 60% maior que o conjunto das internações ocorridas em todo período no município de João Pessoa, que apresenta a maior densidade demográfica da Paraíba.

De acordo com o censo 2010 do IBGE, as cidades de Araruna e Uiraúna ocupam respectivamente a 33^a e a 55^a posições em número de habitantes na Paraíba.

Considerando as outras variáveis, o volume de chuva e os focos do mosquito, só foi possível analisar a variável correspondente às chuvas nas cidades. Porque a busca dos dados sobre os focos do *Aedes aegypti* não foi bem-sucedida para estruturar a análise correspondente ao período de estudo estabelecido no escopo da pesquisa.

Os poucos dados obtidos sobre os focos do vetor nos municípios da Paraíba foram coletados de arquivos em formato PDF divulgados pelo Levantamento Rápido de Índices de Infestação pelo *Aedes aegypti* (LIRAA). Assim, essa análise foi descartada por causa da limitação em se obter os dados referentes a todos os meses de 2014 a 2018 nos municípios de Campina Grande, Araruna, Uiraúna e João Pessoa.

Já em relação aos dados do volume de chuva no período de 2014 a 2018, estes foram coletados através do portal da AESA. E desta forma verificou-se que o volume de chuva nos municípios de Araruna, Uiraúna e Campina Grande concentrou-se numa média de 500 mm a 750 mm nos anos observados.

Exceto João Pessoa que apresentou um volume de chuva maior em relação às cidades citadas.

Tabela 2: Volume de chuva nos municípios mais vulneráveis

| Município | Vol. de Chu. 2014* | Vol. de Chu. 2015* | Vol. de Chu. 2016* | Vol. de Chu. 2017* | Vol. de Chu. 2018* |
|----------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Campina Grande | 705,5 | 595,5 | 544,3 | 607,0 | 679,1 |
| Araruna | 755,2 | 684,0 | 696,3 | 681,0 | 697,7 |
| Uiraúna | 624,2 | 505,3 | 501,5 | 592,8 | 616,4 |
| João Pessoa | 1488,6 | 1322,2 | 1534,7 | 2078,0 | 1525,4 |

*Média em mm calculada no ano (AESA)

Fonte: A autora (2019)

Na Tabela 2 apresentada é possível notar que o volume de chuva nas cidades de Campina Grande, Araruna e Uiraúna apresenta valores aproximados no período restrito. Isto dificulta definir a relação entre o volume de chuva e o número de internações registradas em cada uma delas.

Destaca-se assim, a necessidade de realizar uma análise mais específica, utilizando um escopo maior de dados de chuva, podendo ser dados mensais ou semanais da variação volumétrica. Para que dessa forma seja possível identificar como o volume das chuvas em uma área pode ser relacionado com as de internações apresentadas.

Além disso, considera-se importante analisar as internações em relação a outros fatores de influência em um período maior que cinco anos de registros, como foi adotado neste estudo.

5 CONSIDERAÇÕES FINAIS

A pesquisa teve como objetivo mapear os casos de dengue no Estado da Paraíba através das internações hospitalares registradas no período de 2014 a 2018. Com a finalidade de identificar os municípios mais afetados pela doença e definir o padrão de comportamento dos casos.

A análise exploratória dos dados permitiu que os objetivos da pesquisa fossem atingidos, com ressalva a um dos específicos, possibilitando a produção de conhecimento relevante sobre o contexto da doença dengue na Paraíba.

5.1 Sumário da Pesquisa

Tendo em vista alcançar os objetivos desta pesquisa, algumas questões foram levantadas para serem respondidas no decorrer do trabalho. A análise exploratória dos dados auxiliou a responder às questões de pesquisa, com exceção a um dos objetivos específicos.

Com relação à questão de pesquisa **QSP1**: Quais foram os municípios com os maiores registros de internações por dengue? Percebeu-se através da distribuição geográfica dos casos, que os maiores números de internações se concentraram nos municípios de Campina Grande, Araruna, Uiraúna e João Pessoa. Com destaque para a cidade de Uiraúna, que em cada ano da análise se destacou como um dos cinco municípios mais afetados no Estado.

Porém existe um conjunto de municípios localizados mais a noroeste da Paraíba, sendo eles Belém do Brejo do Cruz (BBC), Brejo do Cruz (BC), Catolé do Rocha (CR), Brejo dos Santos (BDS), Bom Sucesso (BS), Jericó (JE), Riacho dos Cavalos (RC) e São Bento (SB), que apresentaram um agrupamento de casos nos anos de 2014, 2015 e 2016. Esta concentração pode ser resultante da propagação do mosquito vetor de um município para outro, ou mesmo de outros fatores, possivelmente socioeconômicos que apresentem características semelhantes entre as cidades.

A respeito da questão de pesquisa **QSP2**: Qual a sazonalidade dos casos no período observado? Acerca da sazonalidade dos casos de dengue na Paraíba, foi verificado que o maior número de internações por dengue tende a ocorrer no período de março a junho, podendo essa característica estar relacionada com os fatores climatológicos do Estado nessa época do ano. Esse padrão de meses também foi observado na pesquisa de Mendes Neto (2019), na qual observou-se que o período sazonal da dengue na cidade de Patos tende a ocorrer nos meses de março a julho, em razão das variações da precipitação, temperatura e umidade de maneira geral.

Ressalta-se também que ao verificar a distribuição das internações no período de tempo analisado, observou-se que o padrão da dengue na Paraíba não coincide com o contexto observado no Brasil. Apesar do ano de 2016 ser um dos períodos mais críticos tanto no cenário nacional, quanto na Paraíba individualmente, identificou-se que no país houve uma destacável diminuição no número de internações em 2017 e 2018. Enquanto na Paraíba os casos aumentaram significativamente em 2018 quando comparados ao ano de 2017.

A questão **QSP3**: Qual a relação entre o número de internações e a densidade demográfica dos municípios mais afetados? Foi respondida a partir dos resultados obtidos na questão QSP1. Com o reconhecimento dos municípios de Campina Grande, Araruna, Uiraúna e João Pessoa como os mais afetados pela dengue na Paraíba, buscou-se identificar uma relação entre o número de internações e a densidade demográfica de cada cidade, porém tornou-se difícil definir um padrão correspondente a este fator.

Uma resposta negativa foi obtida ao questionar se as cidades mais populosas do Estado apresentaram os maiores registros de internações por dengue. Pois a cidade de João Pessoa, que é o município mais populoso da Paraíba apresentou o menor número de casos quando comparado aos das cidades de Campina Grande, Araruna e Uiraúna. Principalmente no caso de Araruna e Uiraúna que apresentam a densidade demográfica muito mais baixa que a da capital paraibana.

Além da densidade demográfica, buscou-se definir na questão QSP4 relação entre o número de internações e o volume de chuva nas cidades destacadas.

QSP4: Qual a relação entre o número de internações e o volume de chuvas das áreas mais vulneráveis? Não foi possível caracterizar uma relação aparente entre o volume de chuva e o número de internações nas cidades mais afetadas, mas foi identificado que o volume de chuva nos anos de 2014 a 2018 permaneceu constante nas cidades de Campina Grande, Araruna e Uiraúna, com variações na faixa de 500mm e 750mm.

É importante ressaltar que a limitação para definir essa relação foi decorrente do processo de análise ter sido realizado apenas com os dados das médias anuais de chuva. Considera-se que os resultados seriam mais esclarecedores se uma análise fosse realizada com dados mensais ou semanais das chuvas em relação às internações.

Em referência a questão **QSP5**: Qual a relação entre o número de internações e número de focos do mosquito vetor identificados nas regiões? Não foi possível definir resultados acerca dessa questão por causa da limitação no conjunto de dados. Os dados correspondentes ao foco do mosquito *Aedes aegypti* nos municípios da Paraíba não foram obtidos em completude para elaborar uma análise exploratória no período completo de 2014 a 2018.

Apesar da limitação em um dos objetivos específicos, considera-se que os resultados obtidos na pesquisa são relevantes para o direcionamento de novos estudos sobre a doença dengue, assim como podem ser utilizados de suporte para ações de controle e combate dos surtos de dengue no Estado da Paraíba, principalmente nas áreas mais afetadas pela doença.

5.2 Limitações da Pesquisa

No desenvolvimento do trabalho, um dos objetivos específicos não foi atingido em decorrência da limitação no conjunto de dados.

Considerando os dados do volume de chuva nos municípios da Paraíba, a maior dificuldade para definir uma relação com as internações foi que ao avaliar os dados por média anual tornou-se difícil compreender como esse fator influenciou os casos de dengue nas regiões. Portanto, se destaca a necessidade de verificar essa relação considerando dados correspondentes a períodos menores, meses ou semanas, para a variação volumétrica de chuvas.

Com relação ao foco do mosquito vetor, a maior limitação foi decorrente principalmente da ausência de dados acessíveis sobre as verificações em cada município. A busca dos dados correspondentes ao foco do *Aedes aegypti* não foi bem-sucedida. Em consequência disto, não foi possível distribuir os dados e realizar uma análise exploratória com este fator, conforme a definição do objetivo.

5.3 Trabalhos Futuros

Este estudo pode ser ampliado em trabalhos futuros, considerando a relevância do problema da dengue. Sugere-se a inserção de novos dados ao conjunto⁵ já tabulados nesta pesquisa, assim como o desenvolvimento de análises em relação a outros fatores que podem influenciar à dengue no Estado.

Destaca-se ainda a pretensão de verificar especificamente o desenvolvimento da dengue nos municípios mais afetados que foram apresentados neste trabalho. As seguintes questões podem nortear o desenvolvimento dos trabalhos futuros:

- Qual a relação entre os casos de dengue e variáveis socioeconômicas nos municípios de Campina Grande, Araruna, Uiraúna e João Pessoa?

⁵ O conjunto de dados que foi tabulado está disponível no repositório do GitHub: <https://github.com/KeilaLucas/git>.

- Qual a relação entre a variação climática e os casos de dengue registrados na Paraíba?
- Qual o comportamento dos casos de dengue no Estado da Paraíba em um período de 10 anos de registros?

Ainda com relação a expansão da pesquisa, deseja-se desenvolver um modelo preditivo de software capaz de antecipar o número de casos de dengue utilizando os dados das internações hospitalares associados aos dados de fatores socioeconômicos e/ou climatológicos. Visando-se a antecipação de um possível surto da doença nas regiões da Paraíba.

REFERÊNCIAS

- ANTUNES, J. L. F.; CARDOSO, M. R. A. Uso da análise de séries temporais em estudos epidemiológico. **Epidemiol. Serv. Saúde**, Brasília, 24(3):565-576, 2015. doi: 10.5123/S1679-49742015000300024
- BRASIL. ANS. **O que é uma AIH - Autorização de Internação Hospitalar? (RESSARCIMENTO AO SUS)**. 2019a. Disponível em: http://www.ans.gov.br/aans/index.php?option=com_centraldeatendimento&view=operadora&resposta=1559&historico=19798790. Acesso em: 15 mai. 2019.
- BRASIL. Ministério da Saúde. **Dengue: causas, sintomas, tratamento e prevenção**. 2019b. Disponível em: <http://portalms.saude.gov.br/saude-de-a-z/dengue#alerta>. Acesso em: 28 abr. 2019.
- BRASIL. Ministério da Saúde. **Dengue: diagnóstico e manejo clínico adulto e criança**. Secretaria de Vigilância em Saúde - Departamento de Vigilância das Doenças Transmissíveis. 5. ed. Brasília: Ministério da Saúde, 58 p. 2016.
- BRASIL. Ministério da Saúde. **Internações Hospitalares do SUS por local de internação: Notas Técnicas**. 2019c. Disponível em: <http://tabnet.datasus.gov.br/cgi/sih/rxdescr.htm>. Acesso em: 20 mai. 2019.
- BRASIL. Ministério da Saúde. **Ministério da Saúde alerta para aumento de 149% dos casos de dengue no país**. 2019d. Disponível em: <http://portalms.saude.gov.br/noticias/agencia-saude/45257-ministerio-da-saude-alerta-para-aumento-de-149-dos-casos-de-dengue-no-pais>. Acesso em: 28 abr. 2019.
- BRASIL. Ministério da Saúde. **Parâmetros ACE - Municípios**. 2018. Disponível em: <http://portalms.saude.gov.br/noticias/svs/42279-parametros-ace-municipios>. Acesso em: 28 abr. 2019.
- BRASIL. Ministério da Saúde. **SIH – Sistema de Informação Hospitalar do SUS: Manual Técnico Operacional do Sistema**. Brasília, 2017.
- CENTERS FOR DISEASE CONTROL AND PREVENTION (CDC). **Ciclo de vida do mosquito Aedes aegypti**. Disponível em: https://stacks.cdc.gov/view/cdc/41424/cdc_41424_DS1.pdf. Acesso em: 10 abr. 2019.
- CHIAVEGATTO FILHO, A. D. P. Uso de big data em saúde no Brasil: perspectivas para um futuro próximo. **Epidemiol. Serv. Saúde** [online]. 2015, vol.24, n.2, p.325-332. ISSN 1679-4974. <http://dx.doi.org/10.5123/S1679-49742015000200015>.
- CUNHA, R. C. da. **Estudo de confiabilidade dos dados do Sistema de Informações Hospitalares do Sistema Único de Saúde (SIH/SUS) para internações por condições sensíveis à atenção primária**. 2013. Dissertação (Mestrado em Saúde Comunitária) - Universidade Federal da Bahia, Salvador, 2013.
- DANTAS, E. R. G.; PATRÍCIO JUNIOR, J. C. A.; LIMA, D. S. de; AZEVEDO, R. R. de. **O Uso da Descoberta de Conhecimento em Base de Dados para Apoiar a Tomada de**

Decisões. João Pessoa, 2008. Disponível em:

http://www.aedb.br/seget/artigos08/331_331_Artigo_SEGET_EJDR_Versao_Final_010808.pdf. Acesso em: 18 ago. 2019.

DAVI, C. C. M.; PASTOR, A.; OLIVEIRA, T.; NETO, F. B. L.; BRAGA-NETO, U.; BIGHAM, A.; BAMSHAD, M.; MARQUES, E. T. A.; ACIOLI-SANTOS, B. **Severe Dengue Prognosis Using Human Genome Data and Machine Learning.** In: IEEE Transactions on Biomedical Engineering. 2019. doi: 10.1109/TBME.2019.2897285.

DEITEL, H. M.; DEITEL, P. J. **Java: como programar.** PEARSON, ed. 10, 2016.

DHARMAWARDANA, K.G.S.; LOKUGE, J.N.; DASSANAYAKE, P.S.B.; SIRISENA, M.L.; FERNANDO, M.L.; PERERA, A.S.; LOKANATHAN, S. **Predictive model for the dengue incidences in Sri Lanka using mobile network big data.** In: IEEE International Conference on Industrial and Information Systems (ICIIS), Peradeniya, 2017, p. 1-6. doi: 10.1109/ICIINFS.2017.8300381.

Environmental Systems Research Institute, Inc. **ESRI Shapefile Technical Description.** EUA. 1998

FAYYAD, U. M.; Piatetsky Shapiro, G.; Smyth, P.; Uthurusamy, R. **Advances in Knowledge Discovery and Data Mining.** AAAIPress, The Mit Press. 1996.

FOGAÇA, T. K.; MENDONÇA, F. de A. DENGUE NAS AMÉRICAS: Distribuição Espacial e Circulação Viral (1995-2014). **Hygeia**, v. 13, n. 24, p. 175 - 188, 22 jun. 2017.

FONSECA, J. J. S. **Metodologia da pesquisa científica.** Fortaleza: UEC, 2002.

GERHARDT, T. E.; SILVEIRA, D. T. **Métodos de Pesquisa.** Coordenado pela Universidade Aberta do Brasil – UAB/UFRGS e SEAD/UFRGS. Porto Alegre: Editora da UFRGS, 2009.

GIL, A. C. **Métodos e técnicas de pesquisa social.** 6. ed. São Paulo: Atlas S.A., 2008.

GOUVEIA, L. B. **A linguagem R: um ambiente para explorar dados e aprender com eles.** Conferência Hello World. UFP, Porto Universidade Fernando Pessoa. 2017.

HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques.** Morgan Kaufmann Publishers, San Francisco, CA. 2000

ITANO, F.; SANTOS, S. M. dos. **Tópicos de Estatística utilizando R.** Instituto de Matemática e Estatística - Universidade de São Paulo. 2007. Disponível em: <https://cran.r-project.org/doc/contrib/Itano-descriptive-stats.pdf>. Acesso em: 26 ago. 2019.

LACERDA, D.P.; DRESCH, A.; PROENÇA, A.; ANTUNES JUNIOR, J. A. V. Design Science Research: método de pesquisa para a engenharia de produção. **Gestão & produção**, v.20, n.4, p. 741-761, 2013.

LATORRE, M. R. D. O; CARDOSO, M. R. A. Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos. **Revista Brasileira de Epidemiologia**, v. 4, n.3, p. 145-152, 2001.

MARTIN, A. C.; HENNING, E.; WALTER, O. M. F. C.; KONRATH, A. C. Análise de séries temporais para previsão da evolução do número de automóveis no Município de Joinville. **Revista ESPACIOS**, v. 37, n. 06, p. 29. 2015. Disponível em: <http://www.revistaespacios.com/a16v37n06/16370629.html>. Acesso em: 10 set. 2019.

MATHULAMUTHU, S. S.; ASIRVADAM, V. S.; DASS, S. C.; GILL, B. S. **Cluster based regression model on dengue incidence using dual climate variables**. In: IEEE Conference on Systems, Process and Control (ICSPC), Bandar Hilir, 2016, p. 64-69. doi: 10.1109/SPC.2016.7920705.

MEDRI, W. **Análise Exploratória de Dados**. Centro de Ciências Exatas. Departamento de Estatística, Universidade Estadual de Londrina, Paraná. 2011.

MENDES NETO, J. T. **Utilização da Ciência de dados para identificação da relação entre fatores climáticos e a incidência de casos de dengue no município de Patos - PB**. 2019. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Estadual da Paraíba, Patos, 2019.

NASCIMENTO, L. B. do; OLIVEIRA, P. dos S.; MAGALHÃES, D. de P.; FRANÇA, D. D. da S.; MAGALHÃES, A. L. A.; SILVA, J. B.; SILVA, F. P. A. da; LIMA, D. M. Caracterização dos casos suspeitos de dengue internados na capital do estado de Goiás em 2013: período de grande epidemia. **Epidemiol. Serv. Saúde**, Brasília, v. 24, n. 3, p. 475-484, 2015. doi: 10.5123/S1679-49742015000300013.

OLIVEIRA, T. N. **Análise de dados sobre a população do mosquito Aedes Aegypti em Patos PB: Identificação de fatores que podem contribuir para a incidência do mosquito em algumas áreas**. 2018. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Estadual da Paraíba, Patos, 2018.

ONU BRASIL. **OPAS: países das Américas devem se preparar para possíveis surtos de dengue**. Disponível em: <https://nacoesunidas.org/opas-paises-das-americas-devem-se-preparar-para-possiveis-surtos-de-dengue/>. Acesso em: 29 abr. 2019.

ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS). **CID-10: Classificação Estatística Internacional de Doenças e Problemas relacionados à Saúde**. São Paulo: EdUSP, 2015.

PETRUZALEK, D. READ.DBC - Um Pacote para Importação de Dados do Datasus na Linguagem R. In: XV Congresso Brasileiro de Informática em Saúde, Goiânia, 2016. **Anais [...]**. Goiânia: Universidade Federal de Goiás - Comissão de Governança da Informação em Saúde da UFG (CGIS-UFG), p. 601-606. 2016.

PORTO, F. A. M.; ZIVIANI, A. Ciência de Dados. In: 3o. Seminário de Grandes Desafios da Computação no Brasil, 2014, Rio de Janeiro, RJ. **Anais[...]**. 3º Seminário de Grandes Desafios da Computação, 2014.

RAMALHO, I. L. C.; ARAÚJO, F. M. de C.; CAVALCANTI, L. P. de G.; BRAGA, D. N. M.; PERDIGÃO, A. C. B.; SANTOS, F. B. dos; NOGUEIRA, F. de B.; ESCÓSSIA, K. N. F. da; GUEDES, M. I. F. Dengue 4 in Ceará, Brazil: characterisation of epidemiological and laboratorial aspects and causes of death during the first epidemic in the state. **Mem. Inst.**

Oswaldo Cruz, Rio de Janeiro, v. 113, n. 11, e180320, 2018. doi: 10.1590/0074-02760180320.

RED HAT, Inc. **6 pivotal moments in open source history**. Disponível em: <https://opensource.com/article/18/2/pivotal-moments-history-open-source>. Acesso em: 16 ago. 2019.

REIS, M. M. **Estatística aplicada à administração**. Florianópolis: Departamento de Ciências da Administração, UFSC, 2008.

SAS Institute Inc. **Machine Learning: O que é e qual sua importância**. Disponível em: https://www.sas.com/pt_br/insights/analytics/machine-learning.html. Acesso em: 25 ago. 2019.

SOUZA, K. R.; SANTOS, M. L. R.; GUIMARÃES, I. C. S.; RIBEIRO, G. de S.; SILVA, L. K. Saberes e práticas sobre controle do *Aedes aegypti* por diferentes sujeitos sociais na cidade de Salvador. **Cad. Saúde Pública**, Rio de Janeiro, v. 34, n. 5, e00078017, 2018.

Tronchoni, A. B.; Pretto, C. O.; Rosa, M. A. da; Becon Lemos, F. A. DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS DE EVENTOS DE DESLIGAMENTOS DE EMPRESAS DE DISTRIBUIÇÃO. **Revista Controle & Automação**, v. 21, n. 2, p. 185 - 200, 2010.

VERAS, C. M. T.; MARTINS, M. S. A confiabilidade dos dados nos formulários de Autorização de Internação Hospitalar (AIH). Rio de Janeiro, Brasil. **Cad. Saúde Pública**. v.10, n.3, p. 339-355. 1994.

VERDAL, J. C. R.; FILHO, R. C.; VANZILLOTTA, C.; MACEDO, G. L. de; BOZZA, F. A.; TOSCANO, L., PRATA, A. P.; TANNER, A. C.; MACHADO, F. R. Recomendações para o manejo de pacientes com formas graves de dengue. **Revista brasileira de Terapia Intensiva**, v. 23, nº 2, p. 125-133, São Paulo, 2011.

VIANA, D. V.; IGNOTTI, E. A ocorrência da dengue e variações meteorológicas no Brasil: revisão sistemática. **Revista Brasileira de Epidemiologia**, 16(2): p. 240-56, 2013.

VINCENTI-GONZALEZ, M. F.; TAMI, A.; LIZARAZO, E. F.; GRILLET, M. E. ENSO-driven climate variability promotes periodic major outbreaks of dengue in Venezuela. **SCIENTIFIC REPORTS**, v.8, Article number: 5727, 2018. doi:10.1038/s41598-018-24003-z.

WANDERLEY, L. de S. **Aplicação de técnicas de análise de dados para identificar a relação entre fatores ambientais e socioeconômicos com os casos de dengue em Patos PB**. 2018. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Estadual da Paraíba, Patos, 2018.

WICKHAM, H.; GROLEMUND, G. **R Para Data Science (Português)**. O'Reilly, 2017.

WIENER N. **Extrapolation, interpolation and smoothing of stationary time séries**. Cambridge: MIT Press. 1966.